A tamarisk habitat suitability map for the continental United States

Jeffrey T Morisette[1*], Catherine S Jarnevich[2], Asad Ullah[3], Weijie Cai[4], Jeffrey A Pedelty[1], James E Gentle[4], Thomas J Stohlgren[2], and John L Schnase[1]

This paper presents a national-scale map of habitat suitability for tamarisk (*Tamarix* spp, salt cedar), a high-priority invasive species. We successfully integrate satellite data and tens of thousands of field sampling points through logistic regression modeling to create a habitat suitability map that is 90% accurate. This interagency effort uses field data collected and coordinated through the US Geological Survey and nationwide environmental data layers derived from NASA's MODerate Resolution Imaging Spectroradiometer (MODIS). We demonstrate the use of the map by ranking the 48 continental US states (and the District of Columbia) based on their absolute, as well as proportional, areas of "highly likely" and "moderately likely" habitat for *Tamarix*. The interagency effort and modeling approach presented here could be used to map other harmful species, in the US and globally.

Front Ecol Environ 2006; 4(1): 11–17

Tamarisk (*Tamarix* spp, salt cedar) is an Asian tree/shrub species which is invading riparian zones in the United States (Christensen 1962; Robinson 1965). It alters stream hydrology, increases soil salinity, and degrades habitats for native species. There are substantial costs associated with the eradication or control of tamarisk, with implications for water salvage, wildlife use, and riparian restoration (Shafroth *et al.* 2005). Furthermore, many organizations, from federal agencies to grass-roots citizen coalitions, are concerned with tamarisk invasion. For example, the Secretaries of the Interior and Agriculture have called for a cooperative initiative to control invasive tamarisk (USDoI 2005), highlighting a national interest in setting priorities for tamarisk-related control and restoration efforts. These efforts, in turn, require geospatial information on tamarisk distribution, abundance, and suitable habitat at a national scale.

Here we present a map of tamarisk habitat suitability throughout the continental US. This work builds on recent analysis in the western US, showing the abundance of tamarisk in that region (Friedman *et al.* 2005). Our model, based on positive field locations and absence locations, shows that many low- and mid-elevation waterways in western and central US are vulnerable to tamarisk invasion. The potential habitat for tamarisk goes well beyond areas where it already occurs. Along with providing current distribution dara, this habitat map can help guide containment boundaries, identify priority areas for early detection and rapid response, and monitor control strategies and cost-effectiveness in different states. We consider this mapping effort to be a first approximation for mapping tamarisk habitat at the national level. It will be improved upon as more field data become available, additional continental-scale environmental data layers are constructed and incorporated into the model, and users provide feedback.

The habitat map was constructed by coupling field data with geospatial information derived from satellite imagery. The US Geological Survey (USGS) compiled field data indicating the presence or absence of tamarisk from over 40 datasets and covering 32 148 points. The field data provided sufficient information to both construct and test the model. Two-thirds of the data were used to construct the model and one-third was used to test the results. These data were coupled to remote sensing data from the National Aeronautics and Space Administration's (NASA) Earth Observing System through a logistic regression.

Previous studies have also used remote sensing darasets to predict invasive species. For example, Peterson (2005) estimated cover of invasive grasses using a modeling approach similar to that described here, but for a smaller area with higher resolution data. Several studies have shown a relationship between a remotely-sensed spectral response and tamarisk habitat, but again, these are for smaller areas using higher resolution satellite or airborne data (Everitt *et al.* 1989; Everitt *et al.* 1996; Everitt and DeLoach 1990). The novel aspect of the work presented here is its national scale.

The stepwise logistic regression modeling procedure provided an empirical method to relate field data points to environmental layers derived from remote-sensing data covering the contiguous US. Previous work showing the spectral–temporal signature of tamarisk (Everitt and

[1]*NASA Goddard Space Flight Center, Mail Code 614.5, Greenbelt, MD 20771* *(jeff.morisette@nasa.gov); [2]*Fort Collins Science Center, US Geological Survey, Fort Collins, CO*; [3]*Science System Application Inc, Greenbelt, MD*; [4]*George Mason University, Fairfax, VA*

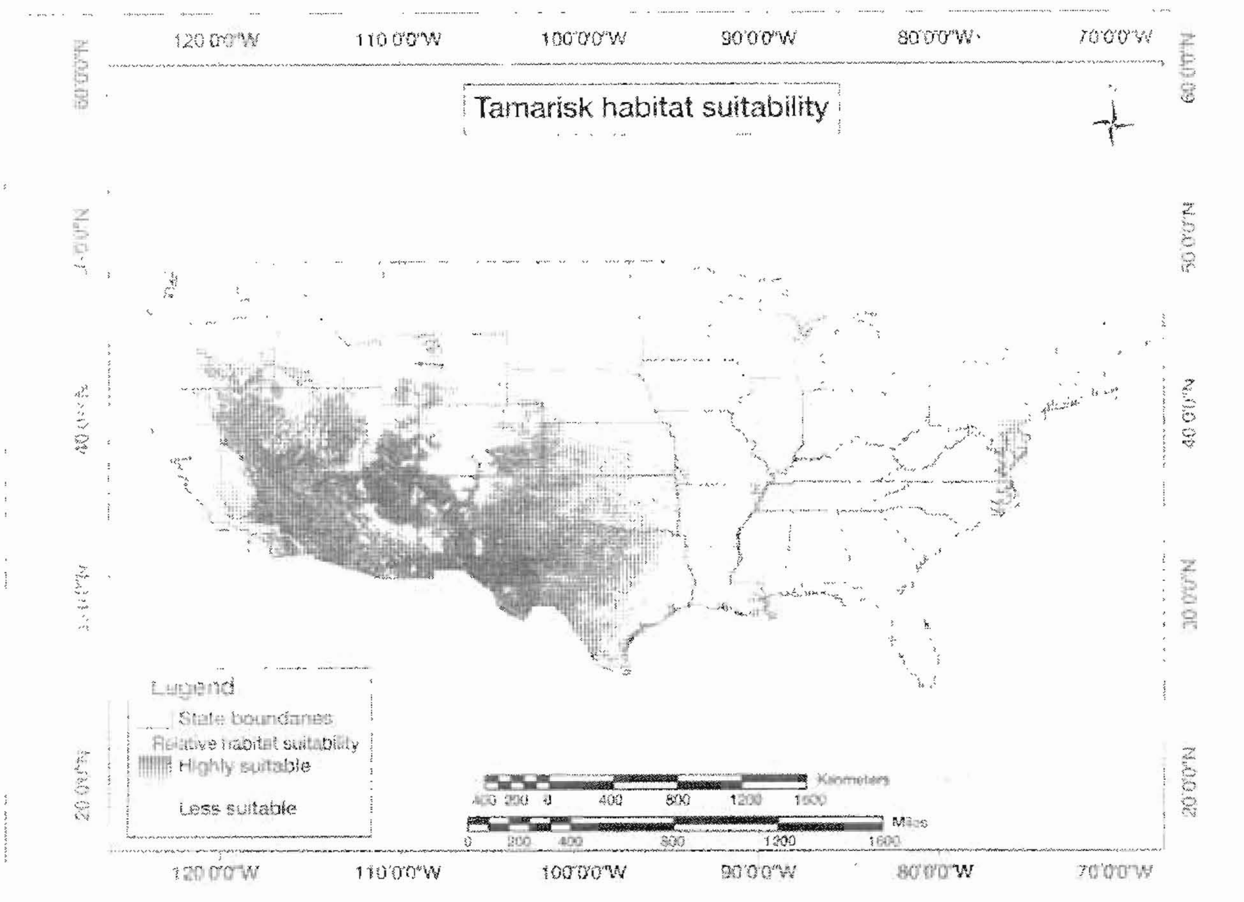

Figure 1. Tamarisk habitat suitability map for the continental US.

DeLoach 1990) led us to exploit the phenology observed in the time series of the MODerate Resolution Imaging Spectroradiometer (MODIS) vegetation index (Huete *et al.* 2002). The model is refined by incorporating land-cover type, also derived from MODIS data (Friedl *et al.* 2002). The stepwise procedure resulted in a highly predictive (90.1%), parsimonious model relating the presence of tamarisk with land-cover type and seasonal variability of vegetation indices.

The logistic regression approach uses the environmental layers to characrerize the habitat of known tamarisk locations as well as those areas with no tamarisk. Areas throughout the continental US exhibiting land-cover and vegetation characteristics similar to locations where tamarisk was observed in the field are associated with a higher metric in the derived map. Areas exhibiting characteristics similar to locations where field data indicated the absence of tamarisk are associated with a lower metric on the map (Figure 1). This metric is then used to classify "highly likely" (areas in the 99th percentile of the map) and "moderately likely" (areas in the 90th percentile) habitat.

Suitable tamarisk habitat is highly variable among states. In Table 1, the two separate columns labeled "rank by fraction" refer to the proportion of either highly or moderately suitable habitat compared to the size of the state. The map and table imply that there is a much greater area of suitable habitat for tamarisk than is currently invaded. (There is no explicit map of all areas that have been invaded, but the number of presence points in Table 1 and the work of Friedman *et al.* [2005] provide an indication of our current understanding.) The Colorado and Rio Grande River basins have experienced heavy infestations, but large areas in the west and southwest are indicated as having suitable habitat for tamarisk and so may be in danger of invasion from adjacent populations. The location and extent of suitable habitat indicares that we may be early in the tamarisk invasion process, or that other factors not measured here are limiting tamarisk spread. Another concern is that hybrids of various tamarisk species may be able to adapt to a wide variety of new habitats on this continent (GISD 2005). Alternatively, strategic containment efforts using biological, chemical, and manual control methods, followed by careful restoration of native species, may slow the spread of tamarisk and associated invasive species. In any case, remote sensing, survey data, and predictive spatial models are important tools for developing efficient and effective containment strategies for non-native species over large areas.

Table 1. Ranking of the lower 48 states (and District of Columbia) by areas of highly suitable (99th percentile) and moderately suitable (90th percentile) tamarisk habitat. For columns reporting area, units = hectares x 1000

State	Number of presence points	Number of absence points	Total area	Area with highly suitable habitat	Rank by area	% of total state area	Rank by fraction	Area with moderately suitable habitat	Rank by area	% of total state area	Rank by fraction
Texas	16	48	68 401	20 598	1	30.11	1	38 657	1	56.52	3
New Mexico	422	0	31 535	3989	3	13.55	2	16 539	3	52.45	5
Nevada	12	1061	28 658	4089	2	12.97	3	18 253	2	63.69	2
Utah	362	697	21 981	3400	4	8.34	4	16 513	4	75.12	1
Arizona	1680	24	29 451	3140	5	8.24	5	16 143	5	54.81	4
California	175	1172	40 787	96	6	0.38	6	14 226	6	34.88	7
Oregon	1	797	25 141	81	7	0.32	7	5664	7	22.53	9
Florida	0	456	14 377	34	9	0.19	8	5004	9	34.80	8
Ohio	0	0	10 670	39	8	0.18	9	5300	8	49.67	6
Colorado	3320	3718	26 962	21	11	0.16	10	3088	11	11.45	13
Wyoming	6	165	25 330	13	14	0.10	11	1611	14	6.36	18
Kansas	6	2	21 289	6	18	0.10	12	873	18	4.10	21
Montana	2	6270	38 134	28	10	0.10	13	3186	10	8.35	16
Idaho	1	1082	21 586	13	13	0.09	14	2582	13	11.96	12
Oklahoma	2	0	18 133	15	12	0.07	15	2688	12	14.82	11
Arkansas	0	22	13 703	6	19	0.06	16	646	19	4.71	19
Indiana	0	0	9427	12	15	0.06	17	1594	15	16.91	10
Alabama	0	185	13 394	6	21	0.05	18	262	21	1.95	25
Illinois	0	0	14 581	10	16	0.04	19	1225	16	8.40	15
Louisiana	0	0	11 816	0	30	0.03	20	100	30	0.84	32
Washington	0	4253	17 363	6	20	0.03	21	555	20	3.20	22
North Carolina	0	2803	12 661	6	17	0.03	22	1169	17	9.23	14
Tennessee	0	814	10 901	3	22	0.02	23	239	22	2.19	24
Virginia	0	0	10 163	1	23	0.00	24	236	23	2.32	23
Mississippi	0	25	12 333	0	27	0.00	25	118	27	0.96	31
Massachusetts	0	0	2081	0	26	0.00	26	143	26	6.85	17
Nebraska	3	39	20 028	0	25	0.00	27	222	25	1.11	29
Vermont	0	0	2487	0	28	0.00	28	114	28	4.59	20
Georgia	0	533	15 175	0	24	0.00	29	225	24	1.48	27
South Carolina	0	755	7986	0	29	0.00	30	102	29	1.27	28
Wisconsin	0	0	14 458	0	32	0.00	31	92	32	0.63	33
West Virginia	0	0	6275	0	31	0.00	32	96	31	1.53	26
Missouri	0	0	18 085	0	33	0.00	33	34	33	0.19	35
North Dakota	0	172	18 339	0	34	0.00	34	32	34	0.17	36
South Dakota	0	262	19 993	0	35	0.00	35	30	35	0.15	37
Minnesota	0	190	21 890	0	36	0.00	36	22	36	0.10	39
New York	0	58	12 200	0	36	0.00	36	6	37	0.05	41
Connecticut	0	0	1288	0	36	0.00	36	6	38	0.49	34
Kentucky	0	0	10 437	0	36	0.00	36	6	39	0.06	40
Michigan	0	138	14 965	0	36	0.00	36	5	40	0.04	44
Iowa	0	63	14 570	0	36	0.00	36	5	41	0.04	43
Pennsylvania	0	80	11 747	0	36	0.00	36	3	42	0.02	45
Rhode Island	0	0	248	0	36	0.00	36	3	43	1.04	30
New Jersey	0	0	1946	0	36	0.00	36	1	44	0.05	42
Delaware	0	0	532	0	36	0.00	36	1	45	0.13	38
Maryland	0	24	2507	0	36	0.00	36	0	46	0.01	47
Maine	0	55	8306	0	36	0.00	36	0	47	0.00	48
New Hampshire	0	0	2398	0	36	0.00	36	0	48	0.00	49
Dist of Columbia	0	77	17	0	36	0.00	36	0	49	0.02	46
Total	**6008**	**26 140**									

There are some caveats related to the map. First, we do not consider sources or pathways for tamarisk introduction. All invasive species require suitable habitat as well as a means of being introduced to the area (ie propagules). Secondly, the map is produced at a spatial resolution of 1 km, a level determined both by the resolution of the input data layers and the practical constraints of preventing the map's file size (~900MB) from becoming too large for access and distribution by a wide range of potential users. Ongoing work is directed at higher resolution, state-level maps and models. At the 1 km resolution, and with the methods employed here, the result is a map of habitat suitability and not the actual presence of tamarisk along watercourses in each 1 km^2 cell, nor the actual susceptible habitat smaller than this resolution (ie narrow riparian zones, springs, etc). It is appropriate to use the map for large-scale summaries (such as those presented in Table 1) or to select focus areas where further analysis with higher resolution imagery and other environmental data layers is justified. Despite these limitations, the results provide a first order approximation of suitable tamarisk habitat and, as such, offer a guide as to which areas across the US should be most closely monitored for tamarisk introduction or spread.

The map is available through the National Institute for Invasive Species Science (NIISS 2005). We welcome and anticipate feedback from its users. In addition, USGS will continue to accumulate tamarisk field data and NASA will continue to explore additional environmental layers that can improve the predictive capacity of the model. The datasets used here were derived from accessible, operational data layers from NASA's MODIS land team (Justice *et al.* 2002). They were readily available fot the study area and their relationship with tamarisk habitat resulted in a good model. Future work could involve additional data layers such as higher resolution remote-sensing datasets, distance to anthropogenic disturbances or to streams or water tables, soils data layers, and climatic variables (such as mean annual temperature as suggested by Friedman *et al.* [2005]). These data layers would have to be available and consistent across the contiguous US and there should be an ecological justification to expect that the additional data layer(s) will improve the prediction of tamarisk habitat. New data layers and additional field points will likely lead to continual improvements in our understanding of tamarisk distributions and suitable habitat.

■ Methods

USGS national tamarisk occurrence data

Field data were collected in three ways. First, beginning in 2001, agencies and organizations, particularly in the state of Colorado, were asked to share information they had collected on the locations of invasive, non-native species. Over 45 disparate datasets were collected and assembled into a single spatial database. This collection effort also involved searching the Internet to locate Geographic Information System (GIS) mapping layers and compiling weed mapping data for several natural areas. The second source of data was unsolicited contributions to the USGS tamarisk mapping project website, T-Map (The *Tamarix* Cooperative Mapping Initiative; www.tamariskmap.org), that was released in April 2004. The final group of data came from fieldwork conducted by our research group and included both presence locations for ramarisk-specific studies beginning in 2003 and presence and absence locations from other vegetation survey field efforts beginning in 1996. Presence locations were extracted from all datasets, including weed mapping data and vegetation plots of all sizes. Absence points were obtained from vegetation survey plots approximating a 30 m^2 grid cell that recorded tamarisk presence. While a measure of tamarisk abundance at a particular site would have provided additional information, in order to maximize the consistency between the disparate datasets we consider only presence and absence here. After all the datasets were combined, vegetation survey data were specifically requested from the VegBank database (http://vegbank.org) to fill in a large data gap for the eastern and northwestern US. The number of presence and absence points from each state is listed in Table 1.

Remotely sensed layers

Constructing a national-level map for the 48 continental states in the US (plus the District of Columbia) requires using the environmental data layers available for that arge area. NASA's MODIS instrument provides almost daily coverage of the globe (Justice *et al.* 2002). The MODIS products described here are the 1 km spatial resolution land-cover product, using the International Geosphere-Biosphere Programme (IGBP) classification system (Friedl *et al.* 2002), the 250 m spatial resolution Normalized Difference Vegetation Index (NDVI), and the Enhanced Vegetation Index (EVI). The EVI was developed to optimize the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring by removing the signal from the background soil and reducing atmospheric influences (Huete *et al.* 2002). To avoid cloud cover and other spurious effects from viewing and illumination angles (Justice *et al.* 2002), the MODIS vegetation index products are generated by compositing daily data every 16 days, resulting in 23 composites per yeat (Huete *et al.* 2002). The MODIS data used here were "Collection 4" data acquired from February 2000 through February 2004, as available through the Land Processes Distributed Active Archive Center (LPDAAC 2005).

Extracting summary values from NDVI and EVI

A discrete Fourier transform was used to extract three summary values from the MODIS NDVI and EVI time series for each pixel (Moody and Johnson 2001). The Fourier transform effectively fits a constant amplitude, yearly sine wave to each pixel, which was uniquely specified by the mean, amplitude, and phase. The ecological interpretation of the mean is the average vegetation greenness from February 2000 to February 2004, while the amplitude of the sine wave describes the average seasonal variability of greenness. The appropriately scaled phase is the average date of peak greenness. This summary method is depicted for one pixel's two-year time series in Figure 2.

Logistic regression

Logistic regression (Hosmer and Lemeshow 2000) was used to associate the binary response of presence or absence of tamarisk with the remote sensing variables (Keane *et al.* 2002). We used the field observation of presence or absence of tamarisk as the dependent variable and considered MODIS land cover and the three summary statistics (mean, amplitude, and phase) from both NDVI and EVI time series as the predictor variables. For the categorical land-cover variable, we used treatment contrasts to set dummy variables with a baseline level of land cover = water. Exploratory data analysis revealed that locations with known tamarisk showed much less absolute difference between the range in NDVI and the range in EVI than did areas without tamarisk (Figure 3). Known ramarisk locations are shown as red crosses on the figure and tend to fall along the line where the seasonal variability in NDVI is equal to the seasonal variability in EVI (shown as a dashed line on Figure 3), while non-tamarisk locations fall off of this line. The difference between the EVI and NDVI MODIS products is an adjustment for the armosphere and soil background (Huete *et al.* 2002). It is probable that the trend of tamarisk growing along the one-to-one line is due to the soil. Tamarisk spreads quickly and is thick enough to covet most soil and will therefore reduce or block any signal from the soil. Conversely, non-tamarisk locations in riparian areas will have either bright sandy or dark wet soils. These will show up as differences in the range in EVI and NDVI in either direction. This theory would match the pattern seen in Figure 3 and led us to consider the absolute difference between the range in EVI and the range in NDVI ($\mathrm{AbsDIFF}_{\mathrm{NDVI-EVI}}$) in the model.

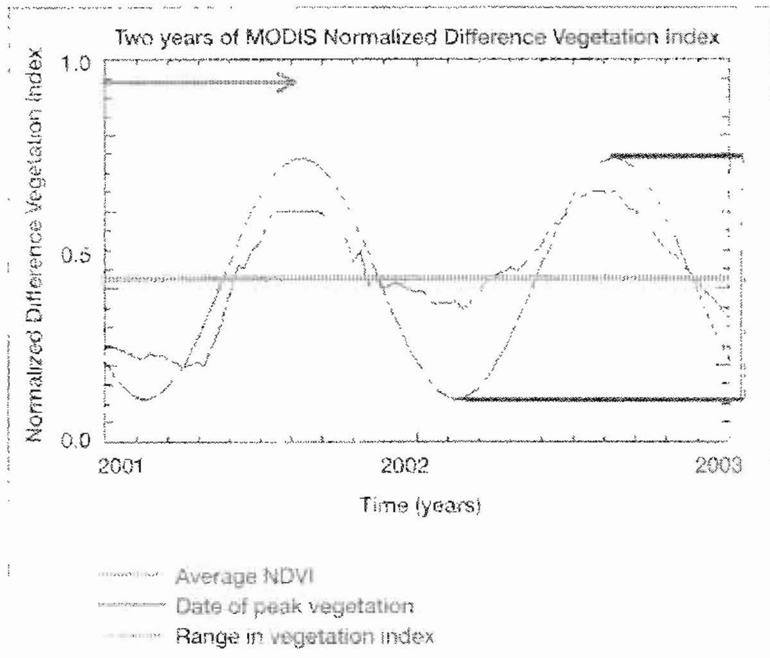

Figure 2. Diagram depicting the summary method applied to each 250 m pixel to derive three metrics from the MODIS vegetation index time series.

The data were split into a training ser to fit the model (using 67% of the data) and a test set to check its accuracy (using the remaining 33% of the data). We maintained a case-control sampling such that the probability of any absence point being included in the sample (P_0) was equal to the probability of any presence point being excluded in the sample (P_1). For the training data both P_0 and P_1 equal 2/3 and for the test data both P_0 and P_1 equal 1/3. It is impossible to know the true proportion of tamarisk habitat in the US and it would be prohibitively difficult at this point to conduct a large enough random sampling to estimate this proportion across the contiguous US. With the data presented here, the proper interpretation of the map is not an absolute probability of the habitat to support tamarisk, but rather a relative ranking of suitable habitat (Keating and Cherry 2004). Thus, the acreage values ranked in Table 1 provide a useful and legitimate interpretation of the logistic tegression results.

We used a forward selection method to find out the variables' entering sequence with regard to their contribution to the modeling. We then used the test dataset to

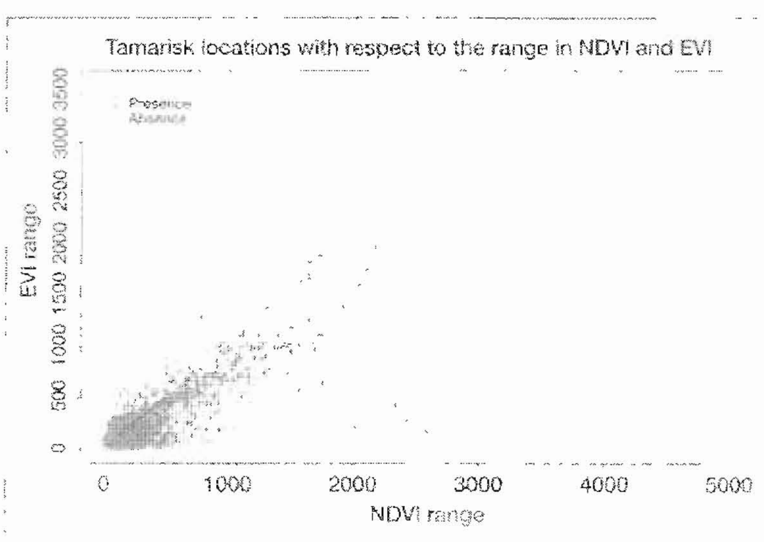

Figure 3. Scatter plot of the range in NDVI vs the range in EVI, with known tamarisk locations shown as red crosses. Non-tamarisk locations are shown as green dots.

Table 2. Logistic regression coefficients associated with the MODIS land cover

Land-cover class	β_i	*Interpretation*
Water	0	(baseline for the categorical variable)
Evergreen needleleaf forest	-2.736	0.065
Evergreen broadleaf forest	-1.118	0.327
Deciduous needleleaf forest	0*	*
Deciduous broadleaf forest	0*	*
Mixed forest	-4.028	0.018
Closed shrub	2.411	11.145
Open shrub	2.511	12.317
Woody savannas	0.426	1.531
Savannas	0.802	2.230
Grasslands	2.071	7.933
Permanent wetlands	0*	*
Croplands	1.583	4.870
Urban and built-up	0*	*
Cropland/natural veg mosaic	-2.922	0.054
Snow and ice	0*	*
Barren and sparsely vegetated areas	1.797	6.032

Values represent the number of times that a particular land-cover class is more likely to support tamarisk than water areas.

Higher values (> 1.0) imply a more suitable land-cover type for tamarisk, while lower values (< 1.0) imply a less suitable habitat for tamarisk.

*These land-cover types have no statistically significant different probabilities of invasion from land-cover type 0.

compare different models and selected the model according to its overall performance. The four criteria used to choose between models were: AUC (area under Receiver-Operating-Characteristic Curve), MSE (mean square error), MAE (mean absolute error), and the proportion of correctly predicted observations with threshold 0.5 (Hosmer and Lemeshow 2000).

The logistic model with best overall performance included the MODIS land-cover variable and the seasonal variability in both NDVI and EVI. The model has the form:

$$\text{habitat suitability} = \frac{\exp(y)}{1 + \exp(y)}$$

where $y = -0.777 - 0.0003281 \times \text{NDVI range} - 0.004735 \times \text{AbsDIFF}_{\text{NDVI-EVI}} + \beta_i$ and β_i depends on the pixel's land-cover type i. Values for β_i are listed in Table 2.

For this model, the AUC = 0.950, MSE = 0.069, MAE = 0.135, and the proportion correct = 0.901. Adding any of the other MODIS vegetation index (either NDVI or EVI) summary variables to the model did not improve any of these criteria. The negative coefficients on both the NDVI range and the $\text{AbsDIFF}_{\text{NDVI-EVI}}$ imply that higher values for these two variables are associated with lower habitat suitability. The interpretation of the MODIS land-cover variable is provided in Table 2.

It is satisfying that such a parsimonious model does a reasonable job fitting this national dataset. The habitat suitability map resulting from the model is appropriate for large-scale analysis (such as the state rankings in Table 1). Further refinement to the national model is being explored with ongoing research at NASA and USGS. Furthermore, the national model and map will be used to guide higher-resolution models at a regional, state-wide level. Finally, the data layers used here are operationally available globally and the modeling presented is fairly general. We therefore believe the approach described in this paper could be used to map other harmful species, both in the US and globally.

Acknowledgments

The research described in this paper was supported by grants through NASA's Terrestrial Ecology program (D Wickland, Program Manager), a NASA Interdisciplinary Science project (W Turner, Program Manager), and the NASA Research, Education, and Applications Solutions Network (REASoN) solicitation (E Sheffner, Program Manager). We thank these NASA program managers for support of our research on integrating NASA data into ecological modeling and invasive species forecasting. Special thanks to J Smith, NASA Goddard Space Flight Center, for his leadership role in the Invasive Species Forecasting System. This analysis was conducted as part of the USGS/NASA Invasive Species Forecasting System project.

References

Christensen EM. 1962. The rate of naturalization of *Tamarix* in Utah. *Am Midl Nat* 68: 51–57.

Everitt JH and DeLoach CJ. 1990. Remote sensing of Chinese tamarisk (*Tamarix chinensis*) and associated vegetation. *Weed Sci* 38: 273–78.

Everitt JH, DeLoach CJ, and Hart WG. 1989. Using remote sensing for detecting brush and weeds on rangelands in the southwestern United States. In: Delfosse ES (Ed). Proceedings of the VII International Symposium on Biological Control of Weeds. 6–11 March 1988, Rome, Italy.

Everitt JH, Escobar DE, Alaniz MA, *et al.* 1996. Using spatial information technologies to map Chinese tamarisk (*Tamarix chinensis*) infestations. *Weed Sci* 44: 194–201.

Friedl MA, McIver DK, Hodges JCF, *et al.* 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens Environ* 83: 287–302.

GISD (Global Invasive Species Database). 2005. *Tamarix ramosissima* (shrub, tree). www.issg.org/database/species/Similar Species.asp?si=72&fr=1&sts=. Viewed 19 August 2005.

Hosmer DW and Lemeshow S. 2000. Applied logistic regression. New York, NY: Wiley-Interscience.

Huete A, Didan K, Miura T, *et al.* 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* 83: 195–213.

Justice CO, Townshend JRG, Vermote EF, *et al.* 2002. An overview of MODIS land data processing and product status. *Remote Sens Environ* **83**: 3–15.

Keane RE, Rollins MG, McNicoll CH, and Parsons RA. 2002. Integrating ecosystem sampling, gradient modeling, remote sensing, and ecosystem simulation to create spatially explicit landscape inventories. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Keating KA and Cherry S. 2004. Use and interpretation of logistics regression in habitat selection studies. *J Wildlife Manage* **68**: 774–89.

LPDAAC (Land Processes Distributed Active Archive Center). 2005. MODIS data products. http://edcdaac.usgs.gov/modis/dataprod.html. Viewed 18 August 2005.

Moody A and Johnson DM. 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sens Environ* **75**: 305–23.

NIISS (National Institute for Invasive Species Science). 2005. National Institute for Invasive Species Science. www.NIISS.org. Viewed 18 August 2005.

Peterson EB. 2005. Estimating cover of an invasive grass (*Bromus tectorum*) using tobit regression and phenology derived from two dates of Landsat ETM+ data. *Int J Remote Sens* **26**: 2491–507.

Robinson TR. 1965. Introduction, spread, and areal extent of saltcedar (Tamarix) in the western states. Washington, DC: US Geological Survey. USGS Professional Paper 491–A.

Shafroth PB, Cleverly JR, Dudley TL, *et al.* 2005. Control of Tamarix in the Western United States: implications for water salvage, wildlife use, and riparian restoration. *Environ Manage* **35**: 231–46.

USDoI (United States Department of Interior). 2005. Press release. www.doi.gov/news/05_News_Releases/050513b. Viewed 18 August 2005.