ÉCOSCIENCE 20 (2): 161-171 (2013)

Presence–absence *versus* invasive status data for modelling potential distribution of invasive plants: Saltcedar in Argentina[1]

Evangelina NATALE[2], Laboratorio de Morfología Vegetal, Departamento de Ciencias Naturales, Universidad Nacional de Río Cuarto, Ruta 36 Km 601 (5800) Río Cuarto, Córdoba, Argentina, and Fundación Conservación y Desarrollo, ConyDes, Sobremonte 1653 (5800) Río Cuarto, Córdoba, Argentina, e-mail: enatale@exa.unrc.edu.ar
Sergio Martin ZALBA, GEKKO, Grupo de Estudios en Conservación y Manejo, Universidad Nacional del Sur, Bahía Blanca, Buenos Aires, Argentina.
Herminda REINOSO, Laboratorio de Morfología Vegetal, Departamento de Ciencias Naturales, Universidad Nacional de Río Cuarto, Ruta 36 Km 601 (5800) Río Cuarto, Córdoba, Argentina.

Abstract: Spontaneous populations of saltcedars are widely distributed in Argentina. The invasive behaviour of the genus has been documented in the USA, Mexico, and Australia, where its presence is associated with significant changes in ecosystem functioning and the structure of natural communities. Previous to this work there were no studies assessing the potential of saltcedars as drivers of ecological change in Argentina. The aim of this work was to assess the potential distribution of saltcedars in the country in order to provide useful information for designing management strategies to reduce the impacts associated with their invasion. Known occurrences of the genus in Argentina were used to predict its potential distribution by applying different distribution models using both presence/absence and presence-only data. The DOMAIN model was the model that performed best once sensitivity and omission errors were taken into account. Our results indicate the severity of the problem of saltcedar in Argentina, with more than three quarters of the total arid and semiarid area vulnerable to invasion. Our results also highlight the need to include information about the status of populations when selecting training points for the development of distribution models.
Keywords: alien species, biological invasion, distribution models, habitat suitability, risk evaluation.

Résumé: Les populations spontanées de tamaris ont une vaste distribution en Argentine. Le comportement envahissant du genre a été documenté aux États-Unis, au Mexique et en Australie, où sa présence est associée à des changements importants dans la structure des communautés naturelles et au fonctionnement des écosystèmes. Aucune étude précédente n'avait évalué le potentiel des tamaris comme moteurs de changements écologiques en Argentine. Le but de la présente étude était d'évaluer la distribution potentielle des tamaris dans ce pays afin d'apporter des informations utiles à la conception de stratégies d'aménagement visant à réduire les impacts de leur invasion. Les présences connues du genre *Tamarix* en Argentine ont été utilisées pour prédire sa distribution potentielle à partir de différents modèles de distribution faisant appel autant aux données de présence/absence qu'à celles de présence seulement. Les résultats ont montré que le modèle DOMAIN était celui qui offrait la meilleure performance en tenant compte de la sensibilité et des erreurs d'omission. Nos résultats mettent en évidence la gravité du problème posé par les tamaris en Argentine avec plus des trois quarts de toute la zone aride et semi aride étant vulnérable à l'invasion. Ces résultats soulignent également la nécessité d'inclure de l'information sur le statut des populations lors de la sélection des données qui serviront au développement des modèles de distribution.
Mots-clés: espèce envahissante, évaluation du risque, habitat adéquat, invasion biologique, modèles de distribution.

Nomenclature: Natale *et al.*, 2008.

Introduction

Invasive alien species are one of the most severe threats to the conservation of biodiversity and a significant cause of economic losses; they also affect human health and welfare (Kadoya *et al.*, 2009). From the arrival of a new species to its establishment and dispersal, the consequences of invasions increase over time; hence, prevention and conservation actions are always the most efficient measures (Ziller & Zalba, 2007). Furthermore, if the species is detected while populations are still relatively small and restricted in space, the chances of controlling its expansion are significantly higher (Inglis *et al.*, 2006). In such cases, modeling the possible distribution of the invader in the new region is very important not only to evaluate its potential impact, but also to detect environmental features that may limit its expansion and that will serve as the basis for designing actions for the prevention, early detection, and early control of new invasion foci (Ward, 2007).

Scientists have long made efforts to forecast the geographical distribution of ecological entities (species,

[1]Rec. 2012-08-14; acc. 2013-03-04.
Associate Editor: Claude Lavoie.
[2]Author for correspondence.
DOI 10.2980/20-2-3571

populations, ecosystems). Species distribution models (SDMs) have contributed to the prediction of species distributions by relating geo-located presence records to environmental variables that influence species' survival and dispersal (Václavík & Meentemeyer, 2009). These models have been used to predict the geographic distributions of plants, microbial pathogens, and arthropods both under present conditions and under the conditions projected in global climate-change scenarios (Mack, Barrett & de Fur, 2002; Guisan & Thuiller, 2005). SDMs are an increasingly important tool for conservation planning and theoretical research on ecological and evolutionary processes (Giovanelli, Haddad & Alexandrino, 2008; Phillips *et al.*, 2009; Mezquida, Rubio & Sánchez-Palomares, 2010).

Used in combination, potential distribution algorithms and habitat suitability maps can assist the design of strategies for the management of invasive species by helping to identify 1) areas where invasive species may already be present (but have not yet been detected) and 2) areas where populations of invasive species may be established in the future (Ward, 2007; Kadoya *et al.*, 2009; Gormley *et al.*, 2011).

A number of algorithms are used to generate SDMs. Different modeling methods may yield very different predictions, depending on the application (Pearson *et al.*, 2006; Randin, Dirnböck & Dullinger, 2006), and this possibility should be taken into account in efforts to generate the most accurate models of species distributions (Phillips & Dudík, 2008).

Potential distribution models pose challenges that may be particularly complex in the study of invasive alien species. On the one hand, spatial distribution can undergo continuous changes (Kadoya *et al.*, 2009). On the other hand, the absence of an invasive species at a given site does not necessarily mean that the habitat conditions at that site are inappropriate; the absence of the species at that particular location may be due to its recent introduction, stochastic events, geographical barriers, and/or dispersal constraints (Araujo & Pearson, 2005; Araujo & Guisan, 2006). Moreover, the local presence of an invasive species could be associated with introduction effort (propagule pressure) and/or with human-induced ecological changes that favour invasion. Hence, new models that use presence-only data or pseudo-absence data generated by the model have been developed (Phillips & Dudík, 2008; Phillips *et al.*, 2009). Because of the urgent need to design preventive and early intervention measures and the shortage of information available on the current distribution of invasive species, other tools, such as regional or national databases (see, for instance, the I3N initiative in the Americas http://i3n.iabin.net/) and other participatory programs, such as mail surveys (Kadoya *et al.*, 2009; Gormley *et al.*, 2011), are particularly useful.

Saltcedars (*Tamarix* spp.) are one of the most successful groups of invasive plants in desert riparian ecosystems in the USA, Australia, and Mexico, where these plants strongly affect species composition and ecosystem processes (De Loach *et al.*, 2000; Australia Weeds Committee, 2004; Hart *et al.*, 2005). The presence of 4 *Tamarix* species has been confirmed in Argentina: *T. gallica*, *T. ramosissima*, *T. chinensis*, and *T. parviflora*. The former 3 grow spontaneously and frequently invade natural and semi natural environments (Natale *et al.*, 2008). The presence of these species in Argentina follows a dispersal pattern similar to that reported for the USA, Australia, and Mexico, colonizing riparian habitats in arid and semi arid continental and coastal areas. *Tamarix ramosissima* has been shown to be the most aggressive saltcedar species, with the highest number of invading populations (Natale *et al.*, 2010). Nevertheless, although arid and semiarid lands cover most of Argentina, to date the magnitude of the saltcedar invasion does not seem to be as severe as that reported in the USA, Mexico, and Australia, probably because the invasion process in Argentina is at an early stage, and may be restrained by a low number of initial invasion foci and/or a low density of dispersal routes (Natale, Zalba & Reinoso, 2012). Hence, the situation in Argentina is suitable for the development and implementation of prevention strategies to reduce the chances of saltcedar expansion.

Considering the successful establishment of saltcedar in the USA, Mexico, and Australia, we hypothesize that these species have the potential to invade most of the arid and semiarid environments in Argentina. If this hypothesis is correct, species of the genus *Tamarix* represent one of the main threats to biodiversity conservation nationwide. To test the hypothesis, presence data of saltcedars in Argentina were used to predict the genus' potential distribution by applying 7 distribution models and comparing their results.

Methods

Collection of saltcedar presence/absence data

The available information on saltcedar occurrences (Natale *et al.*, 2008) was updated by revising herbaria at the Universidad Nacional del Sur (HBBB), the Universidad Nacional de Córdoba (HMBC), the Universidad Nacional de Río Cuarto (RCV Natural Sciences Department and FAV), and the Darwinian Institute (SI). To complement this information, field surveys were conducted covering arid and semi arid regions and areas of adjacent biomes in Argentina, driving along paved and dirt main and secondary routes and local roads. A total of 13 256 km were surveyed between the years 2005 and 2009; 2663 km belonged to the Pampas eco-region (grasslands), 2180 km to the Espinal (temperate deciduous forest), 3734 km to the Monte (dry steppe), 3598 km to the Patagonian desert, and 1588 km to the Chaco (savanna) (Burkart *et al.*, 1996). Saltcedars growing isolated or in groups were spotted by direct observation. We then determined their location using a GPS receiver and recorded the area covered by the trees, their population size structure, evidence (if any) of spontaneous recruitment, the phenological status of the trees, and relevant environmental features (soil texture, salinity, ground water level, distance from watercourse). In addition, following the participatory program method presented by Kadoya *et al.* (2009), electronic surveys were submitted to scientists, technicians, and park-rangers using structured questionnaires. The questionnaire elicited information on 1) population location and size; 2) presence of originally planted individuals;

3) spontaneous recruitment; 4) presence of seedlings more than 20 m from the main population nucleus; 5) seed production; and 6) population structure. We assigned a georeference to the populations cited in the questionnaires and in herbarium records to build up a distribution layer for saltcedar in Argentina.

Of the 297 recorded saltcedar populations (165 from field surveys, 55 from the questionnaires, and 77 from herbaria), 223 that contained sufficient information were classified into 4 categories according to the state of the invasion process (Natale, Zalba & Reinoso, 2012): 1) Contained: populations composed of originally planted individuals, mainly adults (80–100%). 2) Detected in nature: individuals growing in natural and semi natural habitats that had not achieved successful reproduction or did not constitute an invasion at the time of observation. These kinds of populations were mainly composed of young individuals (80–100%). 3) Established: individuals found in natural or semi natural habitats that were successfully spreading, although not widely, to their immediate environment. These populations were 50–80% adults and 20–50% young individuals or seedlings. Finally, 4) Invasive: populations showing extensive dissemination far from the original introduction site and a very heterogeneous structure mostly (20–50% adults, 30–80% young individuals, and 10–40% seedlings). These categories were taken from the IABIN Invasives Information Network (I3N, http://i3n.iabin.net/), and are those employed in the national databases on biological invasions of most countries in the Americas.

For the algorithms using presence/absence data, the sites of saltcedar occurrence were classified into 2 categories: absences corresponded to "contained" populations (47 records), *i.e.*, populations present at a site but showing no evidence of invasion processes, and presences corresponded to populations classified as "detected in nature", "established", or "invasive" (176 records), *i.e.*, foci with confirmed capacity to colonize natural environments. Thus, we minimized the problem of absences that might correspond to sites where the species has not yet been introduced. For complementary purposes, another set of absence data was generated; these records (n = 122), here referred to as "ecological absences", corresponded to sites with environmental characteristics that, as indicated in the available literature on habitat requirements and life history of *Tamarix* spp. (Australia Weeds Committee, 2004; Glenn & Nagler, 2005; Chambers & Hawkins, 2004), would not be suitable for species survival.

Species distribution models

Models using presence/absence data

For the presence/absence model a product developed by Porcasi *et al.* (2005) was used. This product is based on a multi-dimensional map in which each geographic area is characterized by a fixed set of 7 variables. To predict the potential distribution range, Porcasi *et al.* (2005) used the maximum likelihood algorithm, which was run on the ENVI 4.2 platform (Research Systems Inc., Boulder, Colorado, USA) using the supervised classification method. The maximum likelihood algorithm estimates the probability that a pixel belongs to a specific class (presence/absence), with a pre-specified error (0.05). Thus, each pixel is assigned to the class to which it is most likely to belong and those pixels that are below the pre-specified likelihood value are not classified (Chuvieco, 2002). To construct the model the following environmental variables were compiled in a georeferenced environmental database:

1) Precipitation: A latitude–longitude rasterized grid of monthly mean rainfall data over 30 y (1930–1960) with a spatial resolution of 30 min (Leemans & Cramer, 1991; Cramer & Leemans, 2001) (IIASA database, Laxenburg, Austria; available from http://www.daac.ornl.gov).

2) Vegetation Index: A 1982–1992 time series of Normalized Difference Vegetation Index (NDVI) data from the National Oceanic and Atmosphere Administration's Advanced Very High Resolution Radiometer (NOAA/AVHRR) meteorological satellite with a pixel size of 8 × 8 km. NDVI = (Ch2 – Ch1)/(Ch2 + Ch1), where Ch is the AVHRR sensor channel. This index represents greenness but does not distinguish between primary, agricultural, or disturbed habitats.

3) Land Surface Temperature (LST): Another temporal series collected by the AVHRR sensor. LST = Ch4 + 3.33(Ch4 – Ch5) (Price, 1984).

4) Digital Elevation Model (DEM): 1- × 1-km spatial resolution data collected by the AVHRR sensor (provided by the USGS Eros Data Center).

These environmental data were resampled using the nearest neighbour algorithm to adjust the pixel sizes to 1- × 1-km. The resulting 737- × 1037-pixel raw data set covers South America between 13° and 56°S and 33° and 82°W. For this product the authors use the standardized principal component analysis (PCA), because it decomposes the time-series into a sequence of spatial and temporal components that may often be interpreted as corresponding to particular environmental features or events. Typically, the first component indicates the characteristic value of the variable, whereas subsequent components represent change elements of decreasing magnitude. In PCA, spatial variation is registered by the first component and temporal variation is represented by the second and third components. The first 2 components of each environmental variable accounted for more than 99% of the total variance of each series. Thus, both the first and the second components of each environmental variable and DEM were used in the classification analysis. The product generated by principal component analysis is one that is finally used for the presence/absence model.

In order to model saltcedar distribution, 3 data sets were used: 1) presences and confirmed absences; 2) presences and ecological absences, and 3) presences and combined (confirmed and ecological) absences. The Jeffries–Matusita separability test (Richards, 1994) was applied to all 3 data sets to determine similarity or divergence between selected training points. The distance calculated by this index ranges from 0 to 2; if the distance between them is equal to or greater than 1.9, the use of 2 areas or regions as classification types is recommended (Jensen, 1996). This analysis was also performed with the ENVI 4.2 satellite imagery processor.

PRESENCE-ONLY MODELS

These models were developed using 3 different algorithms: BIOCLIM, DOMAIN, and MAXENT. BIOCLIM uses a climate envelope (a rectilinear volume in environmental space) to summarize the climatic data at sites where a species has been recorded. BIOCLIM predicts suitable conditions for a species where values of the climate variables fit within the extreme values determined by the set of known locations. Locations where the values lie within the 5–95th percentile of the climate envelope are traditionally classified as "core" suitability regions (Nix, 1986). DOMAIN (Carpenter, Gillison & Winte, 1993) uses a distance-based method (the Gower metric) to assess new sites according to their environmental similarity to sites of confirmed presence. This model generates an index of habitat suitability on a continuous scale (0–100), in which high scores (*e.g.*, >90) are considered highly suitable. Both the BIOCLIM and DOMAIN modelling methods were implemented using DIVA-GIS software (version 5.2, http://www.diva-gis.org). (Hijmans *et al.*, 2001). MAXENT, developed for use in statistical mechanics, seeks to estimate a target probability distribution by finding the probability distribution of maximum entropy (Phillips, 2006), subject to the constraint that the expected value of each feature under this estimated distribution matches its empirical average. This analysis was performed with the software MAXENT 3.3.1 (Phillips, Dudík & Schapire, 2004).

For presence-only models, environmental data were obtained from WORLDCLIM (version 1.3, http://www.worldclim.org) (Hijmans *et al.*, 2005), taking values of total and average monthly precipitation and minimum and maximum temperature with a spatial resolution of 30 arc seconds (about 1 × 1 km) obtained by interpolation of climate station records from 1950 to 2000. From this climate data, 19 bioclimate variables were derived (Table I), representing a combination of annual trends, seasonality, and extreme environmental conditions.

TABLE I. Climatic variables used in BIOCLIM, DOMAIN, and MAXENT distribution models for *Tamarix* spp. (Hijmans *et al.*, 2005). Variables used in the 12-variable models are shown in bold. STD = Standard deviation.

Climatic variables	Definition	Calculation
Bio1	Annual mean temperature	
Bio2	Mean monthly temperature range	Monthly maximum – mean monthly minimum
Bio3	Isothermality	(Bio2/Bio7)100
Bio4	Temperature seasonality	SDT·100
Bio5	Max temperature of warmest month	
Bio6	Min temperature of coldest month	
Bio7	Annual temperature range	Bio5 – Bio6
Bio8	Mean temperature of wettest quarter	
Bio9	Mean temperature of driest quarter	
Bio10	Mean temperature of warmest quarter	
Bio11	Mean temperature of coldest quarter	
Bio12	Annual precipitation	
Bio13	Precipitation of wettest month	
Bio14	Precipitation of driest month	
Bio15	Precipitation seasonality	Variation coefficient
Bio16	Precipitation of wettest quarter	
Bio17	Precipitation of driest quarter	
Bio18	Precipitation of warmest quarter	
Bio19	Precipitation of coldest quarter	

One of the methodological challenges in the use of these models is to determine the number of variables to be used. To explore the influence of the number of climate variables on model performance, climate data were generated for each species record in DIVA-GIS. PCA was used to assess the similarity between climate variables. Co-linearity was explored through a Pearson correlation matrix, and subsets of variables with a high average correlation (>90%) were reduced to a single variable (Ward, 2007). Both analyses were performed with the software SPSS 11.5.1 (IBM Company, NY, USA).

Unlike the maximum likelihood model, which uses presence/absence data, these models only consider presence data. However, BIOCLIM and DOMAIN require absence data to evaluate model accuracy. For that purpose, the program generated random points across the study area, known as "pseudo-absences". These points were produced in a 1:1 proportion with respect to the presence records and were then contrasted by the model with 25% of the presence records entered initially, which had been reserved for testing.

MAXENT used presence-only data and divided the set of occurrence data into 2 subsets. In this case, 75% of the data were used as training data and the remaining 25% were used for testing (Phillips, Anderson & Schapire, 2006). For BIOCLIM and DOMAIN, the True/False option provided in the DIVA-GIS program was used, with the aim of generating a binary (presence/absence) prediction comparable to the presence/true absence model output. MAXENT's logistic output (ranging from 0 to 1, with 0 indicating absence and 1 indicating presence) was used (Giovanelli, Haddad & Alexandrino, 2008; Sobek-Swant *et al.*, 2012).

The area under the receiver operating characteristic curve (AUC) was also used to examine model performance. AUC measures the ability of a model to discriminate between sites where a species is present and those where it is absent and provides a single measure of overall accuracy that is not dependent on a particular threshold. AUC ranges from 0 to 1, where a score of 1 indicates perfect discrimination and 0.5 implies discrimination that is no better than random (Ward, 2007; Giovanelli, Haddad & Alexandrino, 2008).

Overall, 6 potential distribution maps were generated, 2 per model used, one of them using all the variables (19) and the other one using the variables obtained from PCA and Pearson correlation. In addition, another map was obtained with MAXENT by adding the variable "altitude" to the bioclimate variables (20 variables).

Results

PRESENCE/ABSENCE MODELS

According to the Jeffries–Matusita test, the presence/confirmed absence data resulted in very similar classes (<1.65), indicating that the values of the absence points were very similar to presence data (Pinillas, 1995). Therefore, they were discarded. By contrast, the presence/ecological absence data and presence/combined absence data yielded high separability values (>1.88); hence, both sets were used.

The whole set of presence/ecological absence data showed an AUC value of 0.790, whereas the presence/combined absences data sets yielded an AUC value of 0.860. The presence/ecological absence model showed only a 50% fit to the current saltcedar distribution, whereas the presence/combined absences model included 84.6% of the points (Table II).

Presence-only models

Selection of variables

The first 3 principal components (PC) of the PCA accounted for more than 80% of variation among climatic variables. The first PC accounted for 40% and was mainly represented by temperature variables; the second PC (34.5% of variation) was represented by variables mostly related to precipitation. The third PC accounted for less than 10%; however, it grouped variables related to limiting factors for saltcedar survival (Young, Clements & Harmon, 2004; Natale *et al.*, 2010), such as precipitation of the driest trimester and of the coldest year (Table III).

Table II. Performance of all species distribution models run for *Tamarix* spp. AUC: Area under the Receiver Operating Characteristic curve.

Maximum likelihood algorithm	Model evaluation		
	Sensitivity (%)	AUC	Omission error (%)
Ecological absences	50.0	0.790	42.86
Combined absences	84.6	0.860	5.71
BIOCLIM			
19 variables	75.6	0.782	25.30
12 variables	78.4	0.779	24.00
DOMAIN			
19 variables	98.3	0.850	1.30
12 variables	98.3	0.860	1.30
MAXENT			
20 variables	86.3	0.931	38.00
19 variables	86.3	0.935	28.60
12 variables	73.9	0.943	16.00

Table III. Correlations between climatic variables and principal components from the PCA performed to determine the number of variables to be used in the species distribution models (BIOCLIM, DOMAIN, and MAXENT).

Climatic variables	Principal components			
	1	2	3	4
Bio1	0.814	0.558	–0.016	0.040
Bio2	–0.387	0.753	0.295	0.063
Bio3	–0.675	–0.193	0.613	–0.145
Bio4	0.169	0.853	–0.334	0.280
Bio5	0.619	0.731	–0.153	0.176
Bio6	0.927	–0.127	–0.187	0.064
Bio7	–0.014	0.929	–0.030	0.150
Bio8	0.669	0.513	0.022	–0.435
Bio9	–0.050	–0.009	0.379	0.857
Bio10	0.719	0.674	–0.103	0.106
Bio11	0.897	0.349	0.087	–0.047
Bio12	0.819	–0.498	0.165	–0.008
Bio13	0.842	–0.397	0.322	0.003
Bio14	0.482	–0.761	–0.196	0.118
Bio15	0.206	0.539	0.692	–0.016
Bio16	0.829	–0.399	0.349	0.001
Bio17	0.422	–0.776	–0.353	0.094
Bio18	0.862	–0.108	0.293	–0.205
Bio19	0.192	–0.786	0.098	0.350

Based on the results of the PCA and the Pearson correlation analysis, 12 variables were selected: 1) annual mean temperature, 2) mean monthly temperature range, 3) isothermality, 4) temperature seasonality, 5) minimum temperature of coldest month, 6) mean temperature of wettest quarter, 7) mean temperature of driest quarter, 8) annual precipitation, 9) precipitation seasonality, 10) precipitation of driest quarter, 11) precipitation of warmest quarter, and 12) precipitation of coldest quarter.

BIOCLIM and DOMAIN

With the presence data, the DIVA-GIS software generated 252 presence training points and 412 test points, 374 of which were pseudo-absence points. The BIOCLIM model resulted in an AUC score of 0.782 when run with 19 variables and 0.779 when run with 12 variables. The DOMAIN model showed good predictive power, with an AUC score of 0.850 when run with 19 variables and 0.860 when run with 12 variables (Table II).

Although the number of variables did not influence the fit of the model, it did have a slight influence on prediction of distribution (specificity): the model using 19 variables had a 75.6% fit to the current distribution points, whereas the 12-variable model fit was 78.4%. Hence, both model sensitivity and specificity were higher with the model that used a lower number of variables (12) than with the model that included 19 variables. The results obtained with the DOMAIN algorithm were identical, indicating a good fit of the model for both situations (Table II).

MAXENT

MAXENT used 126 presence training records and 46 test records. The model using 19 variables yielded an AUC score of 0.935, suggesting a very good fit; a similar value was obtained when the model was run with 20 variables (19 bioclimatic variables + altitude). When the model was run with the 12 variables selected, a slightly higher score was obtained, with an AUC of 0.943 (Table II).

Regarding potential distribution, the fit of the prediction to the current distribution of the genus decreased with decreasing number of variables and increasing AUC. The 12-variable model had a 73.9% fit to the current distribution points, whereas the other models had an 86.3% fit (Table II).

Potential distribution of saltcedar in Argentina

The models generated using the maximum likelihood algorithm suggested that the central-western and southern areas of the country would be suitable for the presence of the genus, while the model using the presence/true absence data set predicted the largest area (Figure 1a,b). The latter model was the only one that predicted the presence of the genus in the eastern area of the country (Figure 1b).

The maps generated by BIOCLIM showed minor differences, the 12-variable model being the one that predicted the largest area suitable for the presence of the genus (Figure 2a,b). The 2 maps produced with the DOMAIN algorithm did not show important differences between them, predicting very similar areas. Unlike the other algorithms analyzed, DOMAIN showed new areas suitable for

saltcedar presence, such as the northwest, central-east, and southeast of the country (Figure 3a,b).

The model generated with MAXENT using 12 variables predicted a smaller area with suitable conditions, both in the south and the north of the country. Like DOMAIN, the other models (using 19 and 20 variables) revealed potentially suitable areas in the north of Argentina. The only visible difference was that the model using 20 variables predicted a larger suitable area in the south of the country than the 19-variable model (Figure 4a,c).

Discussion

The results of our study confirm the severity of the problem of saltcedar dissemination in Argentina, with suitable potential distribution areas ranging between 1 452 954 and 1 654 127 km^2, *i.e.*, between 52% and 59.5% of the total surface land of the country and more than three quarters of the arid and semiarid areas in Argentina.

The high similarity found between the presence data and the confirmed absence data used in the maximum likelihood analysis may mean that most of the environments where contained populations are present (*i.e.*, environments where the species is present but has not yet successfully invaded) are suitable for the establishment and spread of the species, but external factors are preventing this from occurring. This would be consistent with the findings of Natale, Zalba, and Reinoso (2012), who reported a close association between contained populations and populations used to provide benefits to humans, *e.g.*, as shade and ornamental trees.

In agreement with the results reported by Evangelista *et al.* (2008), who compared different distribution models for *Tamarix chinensis* in central and southern Utah (USA), we found that the greatest AUC was achieved with MAXENT. Our results are also consistent with data reported for other taxa (Loiselle *et al.*, 2003, Elith *et al.*, 2006; Ward, 2007).

However, overall accuracy as reflected by AUC is not the only important parameter in model selection. It is also important to minimize the risk of false negatives, which is especially relevant in studies of invasive alien species. Assuming that an invasive species will colonize

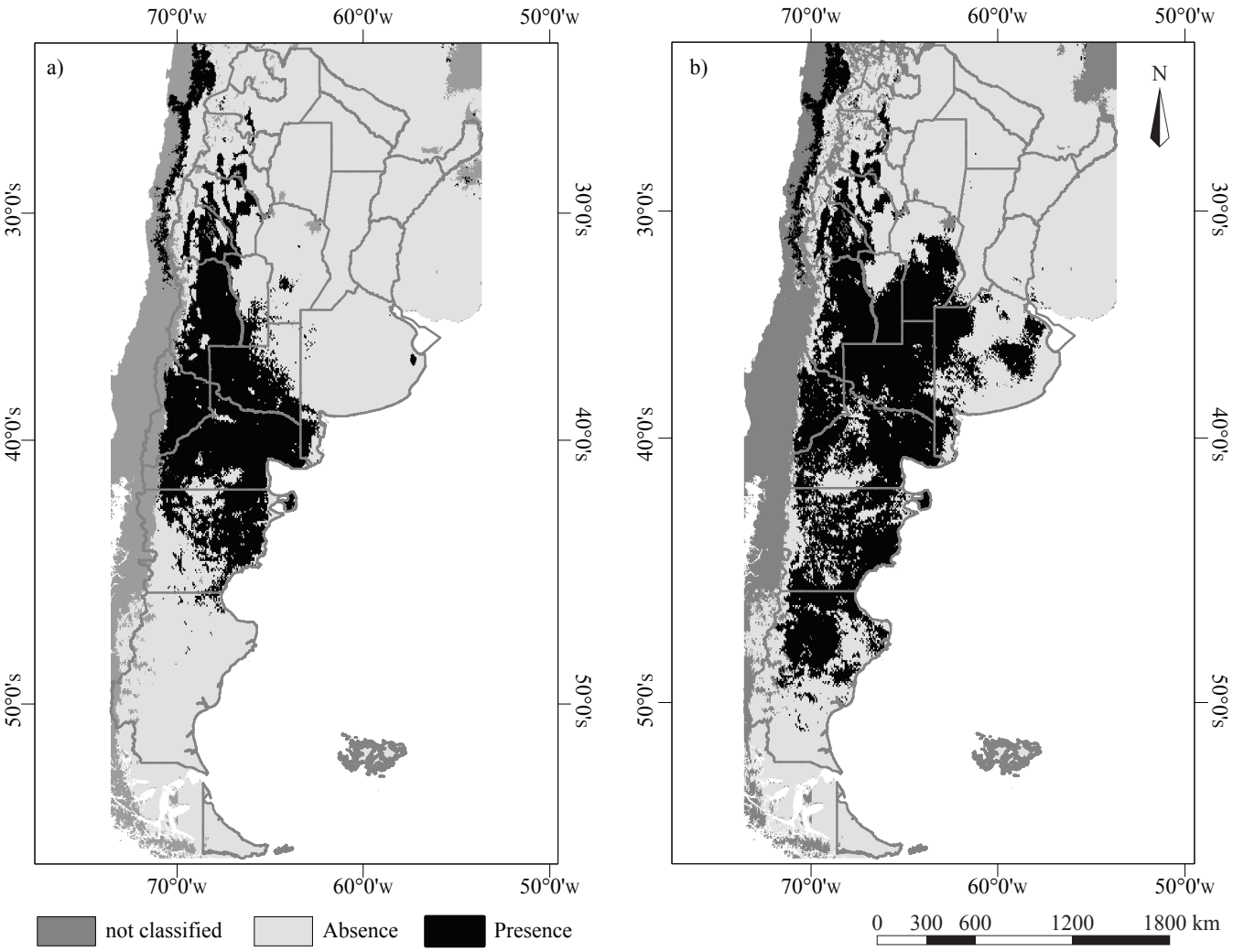

FIGURE 1. Potential distribution of saltcedar using the maximum likelihood algorithm: a) presence/ecological absence data set; b) presence/combined (ecological and confirmed) absence data set.

an area that it will not is preferable to predicting that an invasion will not advance over a region where it actually might. For this reason, preference should be given to models with both a high proportion of correctly predicted presences (*i.e.*, models of high sensitivity) and low omission error (low number of false negatives) (Ward, 2007; Kadoya *et al.*, 2009; Gormley *et al.*, 2011). From this perspective, although MAXENT is the model that performed best in our analysis, its omission error rates (the highest of all models) are unacceptably high. When both of these parameters (AUC and omission error) are taken into account, DOMAIN is the most appropriate model for predicting the potential distribution of *Tamarix* spp. in Argentina. Furthermore, the MAXENT model's results varied increasingly as more variables were incorporated into the analysis. In this respect, our results do not fully agree with those of Elith *et al.* (2006), who reported a general progression of performance (poor to best) from BIOCLIM to DOMAIN to MAXENT. In our study, DOMAIN was consistently one of the best performing models. Because of these differences, we believe that in order to select the most appropriate tool for modelling potential distribution in a particular area, it is important to first compare the results obtained using a variety of different models.

With respect to the use of absence data, Václavík and Meentemeyer (2009) demonstrated that the models incorporating confirmed absence data and factors that restrict species dispersal are the most accurate and efficient. Kadoya *et al.* (2009), however, suggested that including absence data restricts the analysis because those absences may be due to historical restrictions, spread limitations, local extinctions, and/or biological interactions rather than problems related to low habitat suitability for the species. Thus, care should be taken to avoid regarding the mere absence of the species at a site as a valid indicator of environmental unsuitability; the distribution of an invasive species is determined by the combination of environmental factors and dispersal processes (Fukasawa *et al.*, 2009). Accordingly, we believe that the use of sites with contained populations as "absences" is an important step to avoid considering unsuitable those sites that might be suitable, increasing omission errors. This strategy proved to be highly effective in the present work with a maximum

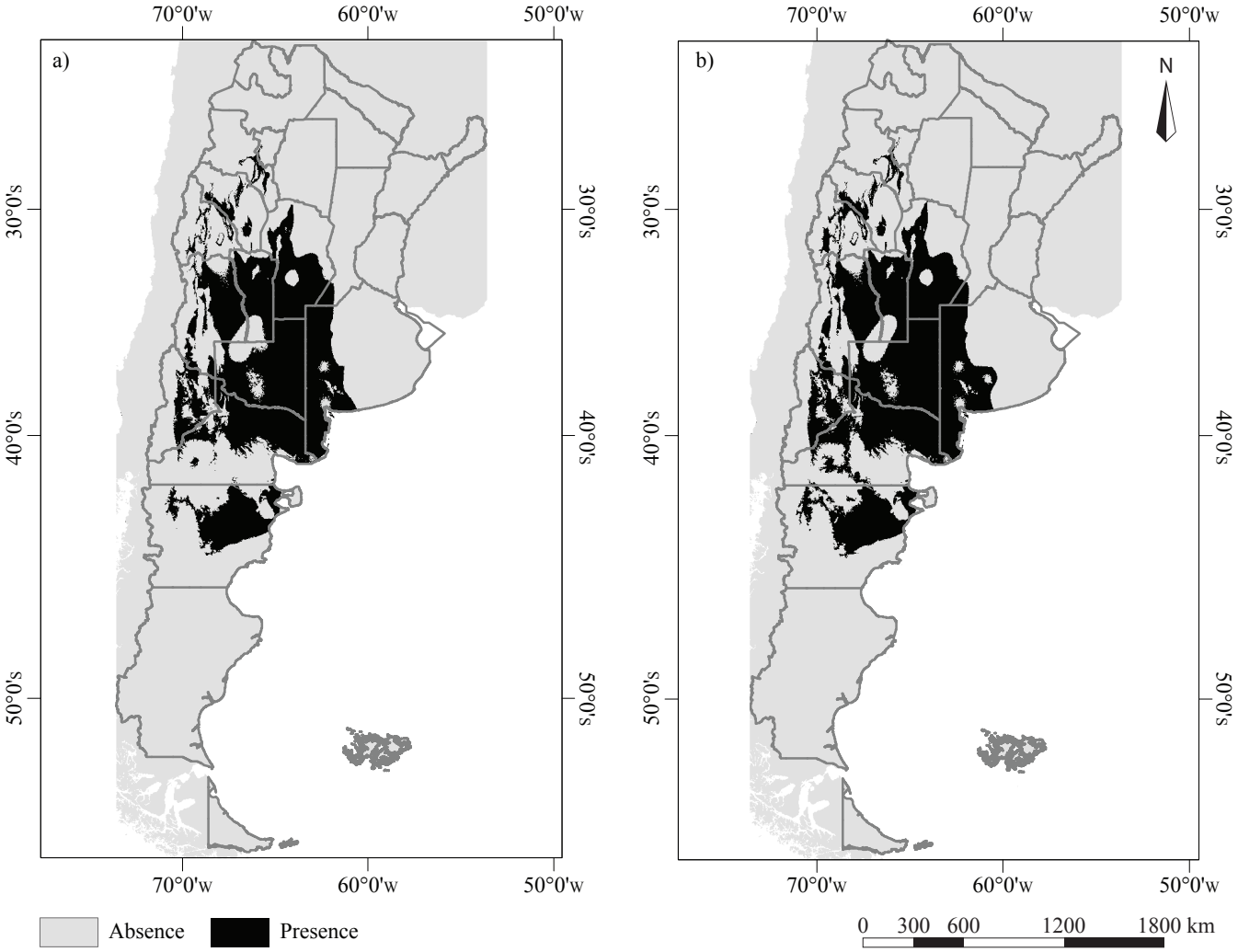

FIGURE 2. Potential distribution of saltcedar using the BIOCLIM algorithm: a) 19 variables; b) 12 variables.

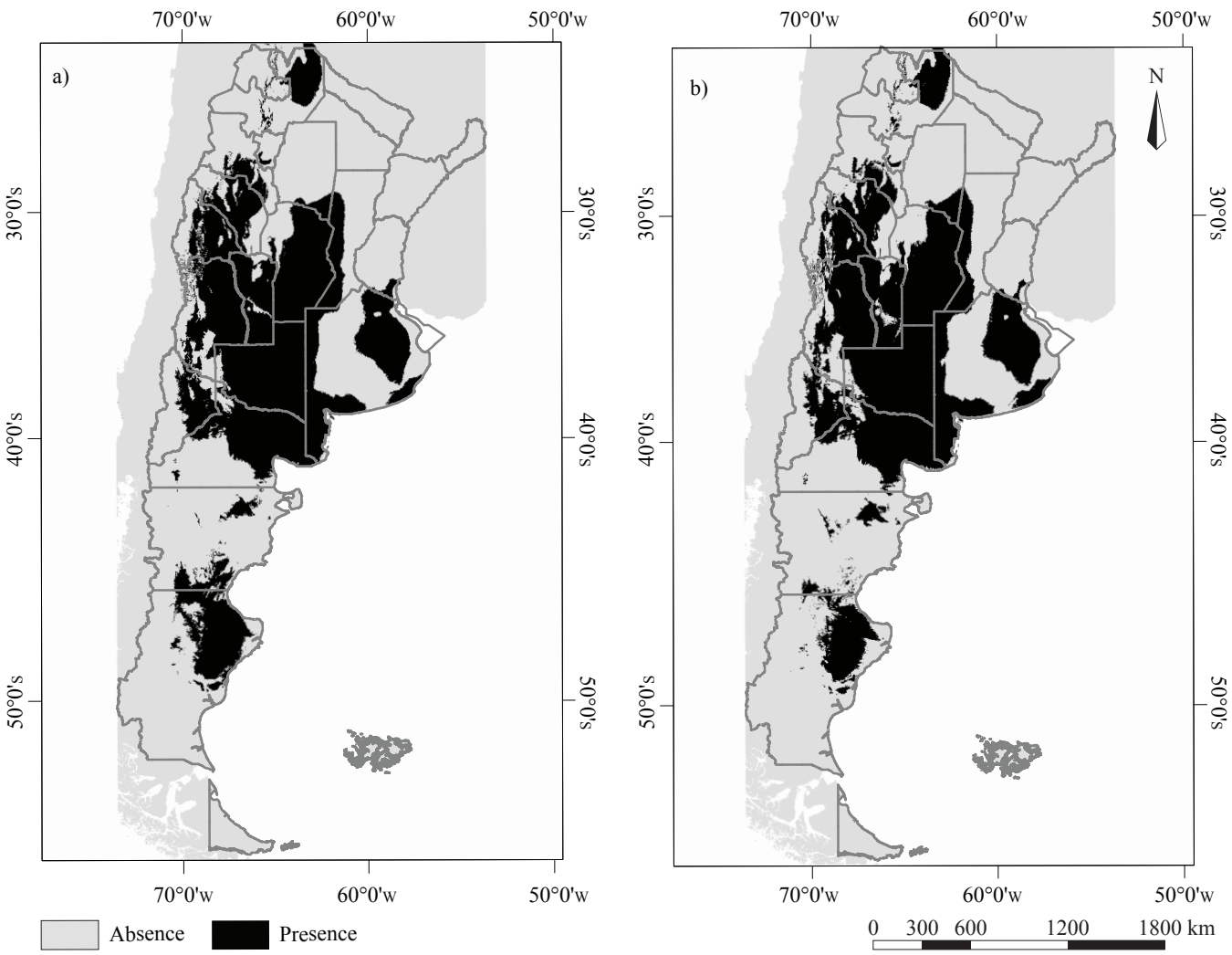

Figure 3. Potential distribution of saltcedar using the DOMAIN algorithm: a) 19 variables; b) 12 variables.

likelihood algorithm calculated using both ecological absences and confirmed absences (contained populations), which resulted in a fit of more than 80% and a very low omission error, very close to that obtained with DOMAIN.

In addition, the mere presence of the species at a site does not imply that such site can be invaded, because the species can survive at a site where environmental conditions prevent their establishment and dissemination (Ziller & Zalba, 2007). Thus, a model considering a mere presence as an occurrence of the species at a site might overestimate the area of potential invasion. To avoid this problem, we considered as presences those sites at which saltcedar populations had shown their potential to colonize natural environments, besides the original plantation sites. We strongly recommend following this criterion (*i.e.*, considering sites with old presence of the species but without evidence of invasion as absences and invaded sites as presences) in data selection so that more robust data are used in invasion modelling.

Finally, it should be noted that, because of the difficulties involved in species identification (especially in information obtained through surveys) the models generated in the present work used a combination of data of the different *Tamarix* species that show invasive behaviour in Argentina. Crawford and Hoagland (2009) stand out the similarity of *Tamarix* species in terms of ecological requirements and also propose potential distribution models at the genus level. However, the several *Tamarix* species exhibit variations in invasive behaviour in different regions worldwide (ARMCANZ, 2000; De Loach *et al.*, 2000; Australia Weeds Committee, 2004; Chambers & Hawkins, 2004; Hart *et al.*, 2005; Whitcraft *et al.*, 2007; Natale *et al.*, 2008). It would be useful to compare the results of the models generated in this work with similar tools applied at the species level when a higher number of records of species distribution are available.

Using participatory surveys proved to be a powerful tool in our study, in agreement with the results reported by Kadoya *et al.* (2009). Structured questionnaires provided *ca* 25% of the data we used to build the models. Regional initiatives aimed at improving and systematizing information about invasive species occurrences and population status should be encouraged as a means to develop sound predictive analysis.

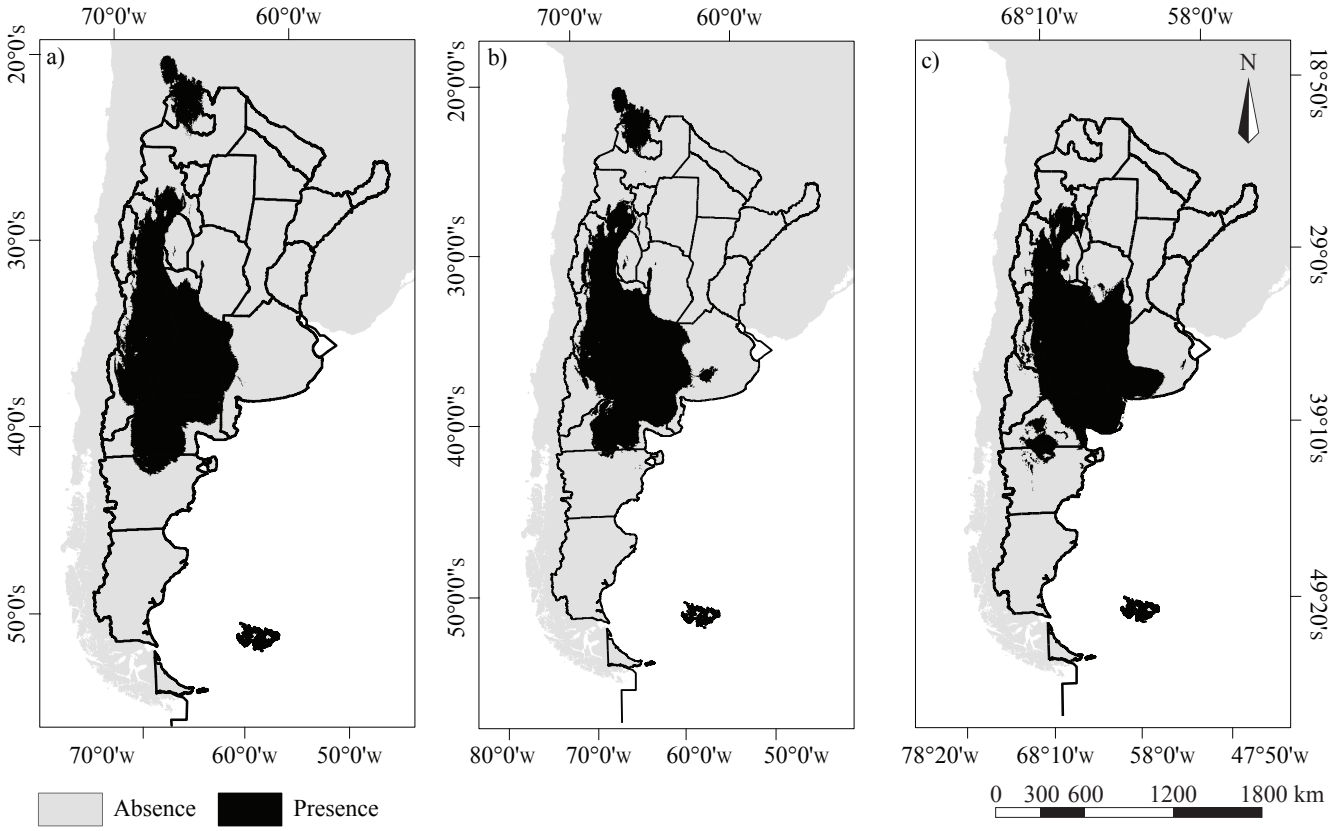

FIGURE 4. Potential distribution of saltcedar using the MAXENT algorithm: a) 19 variables; b) 20 variables; c) 12 variables.

Acknowledgements

Evangelina Natale and Sergio Zalba received financial support from CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina). This work was funded by Secretaria de Ciencia y Técnica de la Universidad Nacional de Río Cuarto and Secretary of Science and Technology, National University of Rio Cuarto (H. Reinoso). The experiments described in this article meet the current laws of Argentina. Authors wish to thank Jorgelina Brasca and Ana Laura Monqaut for reviewing the language of the manuscript.

Literature cited

ARMCANZ (Agriculture, Resources Management Council of Australia and New Zealand, Australian and New Zealand Environment and Conservation Council, and Forestry Ministers), 2000. Weeds of national significance: Athel pine (*Tamarix aphylla*) strategic plan. National Weeds Strategy, Executive Committee, Launceston.

Araujo, M. B. & A. Guisan, 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33: 1677–1688.

Araujo, M. B. & R. G. Pearson, 2005. Equilibrium of species' distributions with climate. Ecography, 28: 693–695.

Australia Weeds Committee, 2004. Weed identification: Athel pine, National Weeds Strategy. Online [URL] http://www.weeds.org.au/cgi-bin/weedident.cgi?tpl=plant.tpl&state=&s=&card=T15 (Accessed on 15 March 2006).

Burkart, R., N. Bárbaro, R. Sánchez & D. Gómez, 1996. Eco-Regiones de la Argentina. Administración de Parques Nacionales, Buenos Aires.

Carpenter, G., A. N. Gillison & J. Winte, 1993. DOMAIN: A flexible modeling procedure of mapping potential distributions of plants, animals. Biodiversity and Conservation, 2: 667–680.

Chambers, N. & T. O. Hawkins, 2004. Invasive Plants of Sonoran Desert: A Field Guide. Sonoran Institute, Environmental Education Exchange and National Birds and Wildlife Foundation (eds.). Online [URL] http://www.sonoran.org/programs/pdfs/invasive_plants_sp.pdf (Accessed on 12 March 2007).

Chuvieco, E., 2002. Teledetección Ambiental: La observación de la Tierra desde el Espacio. 1st Edition. Ariel S.A, Barcelona.

Cramer, W. P. & R. Leemans, 2001. Global 30-Year Mean Monthly Climatology, 1930–1960, Version 2.1 (Cramer and Leemans). Data set from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee. Online [URL] http://www.daac.ornl.gov (Accessed on 10 May 2009).

Crawford, P. H. C. & B. W. Hoagland, 2009. Can herbarium records be used to map alien species invasion and native species expansion over the past 100 years? Journal of Biogeography, 36: 651–661.

De Loach, C. J., R. I. Carruthers, J. E. Lovich, T. L. Dudley & S. D. Smith. Ecological interactions in the biological control of saltcedar (*Tamarix* spp.) in the United States: Towards a new understanding. Faculty Publications (SLS) Paper 72. Online [URL] http://digitalscholarship.unlv.edu/sls_fac_articles/72/ (Accessed on 15 March 2006).

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz & N. E. Zimmermann, 2006. Methods to improve prediction of species' distributions from occurrence data. Ecography, 29: 129–151.

Evangelista, P. H., S. Kumar, T. J. Stohlgren, C. S. Jarnevich, A. W. Crall, J. B. Norman III & D. T. Barnett, 2008. Modelling invasion for a habitat generalist and specialist plant species. Diversity and Distributions, 14: 808–817.

Fukasawa, K., F. Koike, N. Tanaka & K. Otsu, 2009. Predicting future invasion of an invasive alien tree in a Japanese oceanic island by process-based statistical models using recent distribution maps. Ecological Research, 24: 965–975.

Giovanelli, J. G. R., C. F. B. Haddad & J. Alexandrino, 2008. Predicting the potential distribution of the alien invasive American bullfrog (*Lithobates catesbeianus*) in Brazil. Biological Invasion, 10: 585–590.

Glenn, E. P. & P. M. Nagler, 2005. Comparative ecophysiology of *Tamarix ramosissima* and native trees in western U.S. riparian zones. Journal of Arid Environment, 61: 419–446.

Gormley, A. M., D. M. Forsyth, P. Griffioen, M. Lindeman, D. S. L. Ramsey, M. P. Scroggie & L. Woodford, 2011. Using presence-only and presence–absence data to estimate the current and potential distributions of established invasive species. Journal of Applied Ecology, 48: 25–34.

Guisan, A. & W. Thuiller, 2005. Predicting species distribution: Offering more than simple habitat models. Ecological Letters, 8: 993–1009.

Hart, C. H., L. D. White, A. McDonald & Z. Sheng, 2005. Saltcedar control and water salvage on the Pecos River, Texas, 1999–2003. Journal of Environmental Management, 75: 399–409.

Hijmans, R. J., L. Guarino, M. Cruz & E. Rojas, 2001. Computer tools for spatial analysis of plant genetic resources data. 1. DIVA-GIS. Plant Genetic Resources Newsletter, 127: 15–19.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones & A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25: 1965–1978.

Inglis, G. J., H. Hurren, J. Oldman & R. Haskew, 2006. Using habitat suitability index and particle dispersion models for early detection of marine invaders. Ecological Applications, 16: 1377–1390.

Jensen, J. R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. 2nd Edition. Prentice Hall, Englewood Cliffs, New Jersey.

Kadoya, T., S. I. Hiroshi, K. Reina, S. Shin-ichi & W. Izumi, 2009. Using monitoring data gathered by volunteers to predict the potential distribution of the invasive alien bumblebee B*ombus terrestris*. Biological Conservation, 142: 1011–1017.

Leemans, R. & W. P. Cramer, 1991. The IIASA database for mean monthly values of temperature, precipitation and cloudiness of a global terrestrial grid. Research Report RR-91-18. International Institute of Applied Systems Analyses, Laxenburg.

Loiselle, B. A., C. A. Howell, C. H. Graham, J. M. Goerck, T. Brooks, K. G. Smith & P. H. Williams, 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology, 17: 1591–1600.

Mack, R. N., S. Barrett & P. de Fur, 2002. Predicting Invasions of Nonindigenous Plants and Plant Pests. National Academies Press, Washington, DC. Online [URL] http://www.nap.edu/catalog.php?record_id=10259 (Accessed on 15 March 2006).

Mezquida, E. T., A. Rubio & O. Sánchez-Palomares, 2010. Evaluation of the potential index model to predict habitat suitability of forest species: The potential distribution of mountain pine (*Pinus uncinata*) in the Iberian Peninsula. European Journal of Forest Research, 129: 133–140.

Natale, E., S. M. Zalba & H. Reinoso, 2012. Assessing invasion process through pathway and vector analysis: Case of saltcedar (*Tamarix* spp.). Management of Biological Invasions, 3: 37–44.

Natale, E., J. Gaskin, S. M. Zalba, M. Ceballos & H. Reinoso, 2008. Especies del género *Tamarix* (Tamarisco) invadiendo ambientes naturales y seminaturales en Argentina. Boletín de la Sociedad Argentina de Botánica, 43: 137–145.

Natale, E., S. M. Zalba, A. Oggero & H. Reinoso, 2010. Establishment of *Tamarix ramosissima* under different conditions of salinity and water availability: Implications for its management as an invasive species. Journal of Arid Environment, 74: 1399–1407.

Nix, H., 1986. A biogeographic analysis of Australian Elapid snakes. Pages 4–15 in R. Longmore (ed.). Atlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra.

Pearson, R. G., W. Thuiller, M. B. Arau, J. E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson & D. C. Lees, 2006. Model-based uncertainty in species range prediction. Journal of Biogeography, 33: 1704–1711.

Phillips, S., 2006. Una Breve Guía Didáctica sobre Maxent. Online [URL] http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial-in-spanish.doc (Accessed on 4 June 2008).

Phillips, S.J., R. P. Anderson & R. E. Schapire, 2006. A maximum entropy modelling of species geographic distributions. Ecological Modelling, 190: 231–259.

Phillips, S. J. & M. Dudík, 2008. Modelling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography, 31: 161–175.

Phillips, S. J., M. Dudík & R. E. Schapire, 2004. A maximum entropy approach to species distribution modelling appearing. Pages 472–486 in Paper presented at the 21st International Conference on Machine Learning. ACM Press, New York, New York.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick & S. Ferrier, 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications, 19: 181–197.

Pinillas, C., 1995. Elementos de Teledetección. RA-Ma Editorial, Madrid.

Porcasi, X., G. E. Calderón, M. Lamfri, M. Scavuzzo, M. S. Sabattini & J. Polop, 2005. Predictive distribution maps of rodent reservoir species of zoonoses in Southern America. Mastozoología Neotropical, 12: 199–216.

Price, J. C., 1984. Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer. Journal of Geophysics Research, 89: 7231–7237.

Randin, C. F., T. Dirnböck & S. Dullinger, 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography, 33: 1689–1703.

Richards, J. A., 1994. Remote Sensing Digital Image Analysis. Springer-Verlag, Berlin.

Sobek-Swant, S., D. A. Kluza, K. Cuddington & D. B. Lyons, 2012. Potential distribution of emerald ash borer: What can we learn from ecological niche models using Maxent and GARP? Forest Ecology and Management, 281: 23–31.

Václavík, T. S. & R. K. Meentemeyer, 2009. Invasive species distribution modelling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220: 3248–3258.

Ward, D. F., 2007. Modelling the potential geographic distribution of invasive ant species in New Zealand. Biological Invasions, 9: 723–735.

Whitcraft, C. R., D. M. Talley, J. A. Crooks, J. Boland & J. Gaskin, 2007. Invasion of tamarisk (*Tamarix* spp.) in a southern California salt marsh. Biological Invasions, 9: 875–879.

Young, J. A., C. D. Clements & D. Harmon, 2004. Germination of seeds of *Tamarix ramosissima*. Journal of Range Management, 57: 475–481.

Ziller, S. R. & S. M. Zalba, 2007. Proposals to prevent and control exotic invasive species. Natureza & Conservação, 5: 78–85.