A comparison between GARP model and SVM regression to predict invasive species potential distribution: the case of Miconia calvescens on Moorea, French Polynesia

R. Pouteau*, J.-Y. Meyer, R. Taputuarai, B. Stoll

a South Pacific Geosciences Laboratory, University of French Polynesia, BP 6570, 98702 Fa’a’a, French Polynesia - (robin.pouteau, benoit.stoll@upf.pf)
b Research Department of the Government of French Polynesia, BP 20981, Papeete - jean-yves.meyer@recherche.gov.pf
c MaNature, BP 5819, Pirae - rtaputuarai@gmail.com

Abstract - Biological invasion is one of the main drivers for global biodiversity loss. In the islands of French Polynesia, the worst invasive alien species is arguably the small tree Miconia calvescens. The knowledge of miconia distribution and dynamics is critical for monitoring and control efforts. However, classical direct remote sensing methods are vain because miconia invades rainforest understory. In this study, we introduce the possibility to use a Support Vector Machine (SVM) regression for ecological niche modelling and we compare it with the commonly used Genetic Algorithm for Rule-set Production (GARP). Both models integrate several environmental layers extracted from a Digital Elevation Model (DEM) such as elevation, slope, aspect, windwardness and wetness. SVM regression achieves an overall accuracy of 92.5% and outperforms significantly the GARP model. These results may be explained by the ability of SVM regression to integrate heterogeneous data and to be trained on small sets of pixels. We suggest that SVM regression can be used as an effective tool to predict invasive (but also native) species potential distribution.

Keywords: Ecological niche modelling; Digital elevation model (DEM); Support vector machines (SVM); Genetic Algorithm for Rule-set Production (GARP); Invasive species; Miconia calvescens.

1. INTRODUCTION

Invasive alien species are recognized as a primary cause of global biodiversity loss and an agent of global change (Millennium Ecosystem Assessment, 2005). Because of their isolation and small areas, island ecosystems are particularly vulnerable to biological invasions (Loope and Mueller-Dombois, 1989; Paulay, 1994; Kueffer et al. 2010). According to Atherton et al. (2007), biotic invasions represent the first threat to the insular native biota in the “Pacific Ecoregion”. They underline the need for “more information on the distribution and impact of invasive species in sensitive sites and the identification of alien-free habitats” as their first strategic direction.

Advances in the spatial and spectral resolutions of sensors now allow a direct detection at species or community levels (Xie et al. 2008). However, some of the most aggressive weeds invade forest understory with a low or no influence on vegetation spectral response, thus remain undetectable by classical direct remote sensing methods. This kind of problem leads macro-ecologists to develop indirect approaches as well as ecological niche modelling.

In this paper, we propose to compare two ecological niche models to predict the potential distribution of an invasive tree in the island of Moorea, French Polynesia.

2. MATERIAL AND METHODS

2.1 Miconia in French Polynesia

As in most of the small Pacific islands, the native forests of French Polynesia have been invaded by alien plants, accidentally or purposefully introduced by humans (Meyer, 2004). Among the ca. 60 alien plants species considered to be highly invasive (Fourdrigniez and Meyer, 2008), Miconia calvescens DC (Melastomataceae) is probably the worst. It belongs to the world’s 100 worst invasive species in the world (ISSG, 2010). This small tree, commonly 4 to 12 m tall (up to 16 m), and native to rainforests of Central and South America, was first introduced to Tahiti in 1937 as an ornamental plant. It has currently invaded about two-thirds of the forested area on this island (ca. 800 km²) (Meyer, 2009). A relative shade-tolerant species, miconia is commonly found in the understory of native rainforests and cloudforests, thus difficult to detect with classical direct remote sensing methods.

2.2 Study site

On the island of Moorea (140 km² with a highest summit reaching 1,207 m) where miconia has been introduced in the early 1970’s (Meyer, 1993), some distribution maps based of ground surveys were produced in the past (Anonymous, 1990; Meyer, 1993; Schwartz, 1993) but remain incomplete and outdated. The species current distribution area was estimated to be 35 km² i.e. 25% of the total island surface based on intensive field surveys (Meyer 2009; Meyer and Taputuarai, unpub. data).

2.3 Distribution models

Species distribution models need an input of presence and absence ground control pixels and a set of environmental layers. Point localities where miconia occurs or not, respectively called “invaded pixels” and “healthy pixels” hereinafter, were picked up in the field with a Trimble® GeoXH™ handheld GPS. Eighty pixels were used in each set of pixels (presence and absence). For each set of pixels, we devoted 50% (i.e. 40 pixels) to model training and other 50% to model evaluation.

* Corresponding author.
** This work was partially supported by the Research Department of the Government of French Polynesia.
Five environmental layers were used in the species distribution analysis representing parameters that, based on our field experience, we know to influence miconia distribution. These environmental layers were all extracted from a 5m-resolution digital elevation model (DEM) and include elevation, slope, aspect, windwardness and a wetness index (Gessler et al., 2000). The wetness index was used as an index of water drainage, with low wetness index values representing convex positions like mountain crests and with high wetness index values representing concave positions like coves or hillslope bases. It is a function of the slope angle $\beta$ (in radians) and the specific catchment area ($A_s$) expressed as $\text{m}^2$ per unit width orthogonal to the flow direction (1).

$$\text{Wetness index} = \ln\left( \frac{A_s}{\tan \beta} \right)$$  \hspace{1cm} (1)

From this data, we compare the output of two models to map miconia on Moorea.

2.3.1 GARP

Genetic Algorithm for Rule-set Production (GARP) was successfully used for wild species distribution mapping in many studies (Peterson et al., 2003; Roura-Pascual et al., 2004; Loo et al., 2006; LaRosa et al., 2007; Sánchez-Flores, 2007; Ebeling et al., 2008; Roura-Pascual et al., 2008). GARP is a machine learning-based analytical package developed by Stockwell and Noble (1991). More details on the program are available at the URL http://www.nhm.ku.edu/desktopgarp/index.html where the Desktop version used in this study can also be downloaded. Due to unstable traits of the GARP algorithm producing various solutions, we choose to keep the map producing the best accuracy after 100 runs.

2.3.2 SVM

Support Vector Machines (SVM) are arguably one of the most successful machine learning algorithms developed in the last decade. SVM consists in projecting vectors into a high dimension feature space by means of a kernel function then fitting an optimal hyperplane that separates classes using an optimization function (2). For a generic pattern $x$, the corresponding estimated label $y$ is given by (2).

$$y = \text{sign}(f(x)) = \text{sign} \left( \sum_{i=1}^{N} a_i y_i K(x_i, x) + b \right)$$  \hspace{1cm} (2)

Wherein $N$ is the number of training points, the label of the $i$th sample is $y_i$, $b$ is a bias parameter, $K(x_i, x)$ is the chosen kernel function and $a_i$ denotes the Lagrangian multipliers.

SVM is introduced by Vapnik (1998) and extensively described by Burges (1998), Schölkopf and Smola (2002) and Hsu et al. (2009). As suggested by the latter and supported by many other papers, we use the radial basis function (RBF) as the kernel. The equation is (3).

$$K(x_i, x) = \exp \left( \frac{-||x_i - x||^2}{2\sigma^2} \right)$$  \hspace{1cm} (3)

Noise in the data can be accounted for by defining a distance tolerating the data scattering, thus relaxing the decision constraint. This regularization parameter $C$ as well as $\sigma$ is found by cross-validation.

3. RESULTS

SVM regression achieves an overall accuracy (OA; based on the mean of accuracy of both classes “invaded” and “healthy”) of 92.5% and outperforms significantly the GARP model, achieving an OA of 72.5%. Comparing attribution of validation pixels, the GARP model fairly overestimates miconia distribution: 18 healthy pixels are classified as invaded (Table A) while only 5 healthy pixels are labeled as invaded with SVM regression (Table B), which can explain inadequacies of GARP face to SVM regression for mapping miconia on Moorea. Regarding the produced maps (Figure 1), both models gives globally similar distributions localized on non-inhabited highland including slopes and valley bottoms excluding ridges and highest summits. The GARP model produces more homogeneous distribution area while SVM regression seems more sensible to micro-topography which can explain its fairly good results.

| Table A. Confusion matrix of GARP model classification |
|---------------------------------|-------|-------|
| Invaded pixels | 36 | 18 |
| Healthy pixels | 4 | 22 |

| Table B. Confusion matrix of SVM regression classification |
|---------------------------------|-------|-------|
| Invaded pixels | 39 | 5 |
| Healthy pixels | 1 | 35 |

According to the GARP model, miconia potential distribution on the island of Moorea represents 29 km² (21% of the surface of the island) while the SVM regression gives a distribution area of 35 km² (25%).

4. DISCUSSION

The SVM regression model was successfully applied to predict miconia distribution on Moorea.

SVM regression outperforms the GARP model but its accuracy is not 100%. The reasons that may explain this relative inaccuracy include: (i) the SVM algorithm we used is intrinsically not optimal; (ii) miconia invasion is not saturated and some pixels identified as “healthy” in the field might be invaded by miconia in the future (iii) environmental layers are not sufficient to explain the totality of miconia distribution and other variables such as rainfall (data not available on Moorea) or dominant vegetation (work in progress) maps should be integrated.

Our results corroborate the “real” distribution of miconia based on intensive ground surveys. SVM has shown to be a stable algorithm that integrates effectively heterogeneous data and that can be trained on small sets of pixels (Foody, 2006). It may be an excellent candidate to integrate other environmental variables such as dominant vegetation maps in future work. Moreover, SVM regression might be used to map rare endemic species for which, in general, few presence ground control pixels are known.

In this paper, SVM regression was applied to predict the distribution of an invasive plant but it can also be an effective tool for native and/or animal species mapping.

ACKNOWLEDGEMENTS

The authors are grateful to the Government of French Polynesia and its Urbanism Department for providing the DEM data. We also want to thank Marie Fourdrigniez for her help during field surveys and the Moorea Biocode Project for logistic support.
Figure 1. Comparison of miconia potential distribution modeled by GARP (above) and by a SVM regression (below).
REFERENCES


J. Schwartz, Distribution of an invasive introduced plant, Miconia calvescens, on the island of Moorea, French Polynesia. Student Research Paper, The Biology and Geomorphology of Tropical Islands, Fall Semester, University of California at Berkeley, 1993.


