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[a,*], J.-Y. Meyer[b], R. Taputuarai[c], B. Stoll[a]

[a] South Pacific Geosciences Laboratory, University of French Polynesia, BP 6570, 98702 Faa'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 (Millenium 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 purposely 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.

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 β (in radians) and the specific catchment area (As) expressed as m² per unit width orthogonal to the flow direction (1).

$$Wetness\ index = \ln\left(\frac{As}{\tan\beta}\right) \qquad (1)$$

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 algorithm 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 $\mathbf{x}$, the corresponding estimated label $\hat{y}$ is given by (2).

$$\hat{y} = \text{sign}[f(\mathbf{x})] = \text{sign}\left[\sum_{i=1}^{N} a_i.y_i.K(\mathbf{x_i},\mathbf{x}) + b\right] \qquad (2)$$

Wherein N is the number of training points, the label of the ith sample is y_i, b is a bias parameter, $K(\mathbf{x_i},\mathbf{x})$ 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).

$$K(\mathbf{x_i},\mathbf{x}) = \exp-\frac{\|\mathbf{x_i}-\mathbf{x}\|^2}{2\sigma^2} \qquad (3)$$

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 σ 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

GARP	Invaded pixels	Healthy pixels
Invaded pixels	36	18
Healthy pixels	4	22

Table B. Confusion matrix of SVM regression classification

SVM	Invaded pixels	Healthy pixels
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.
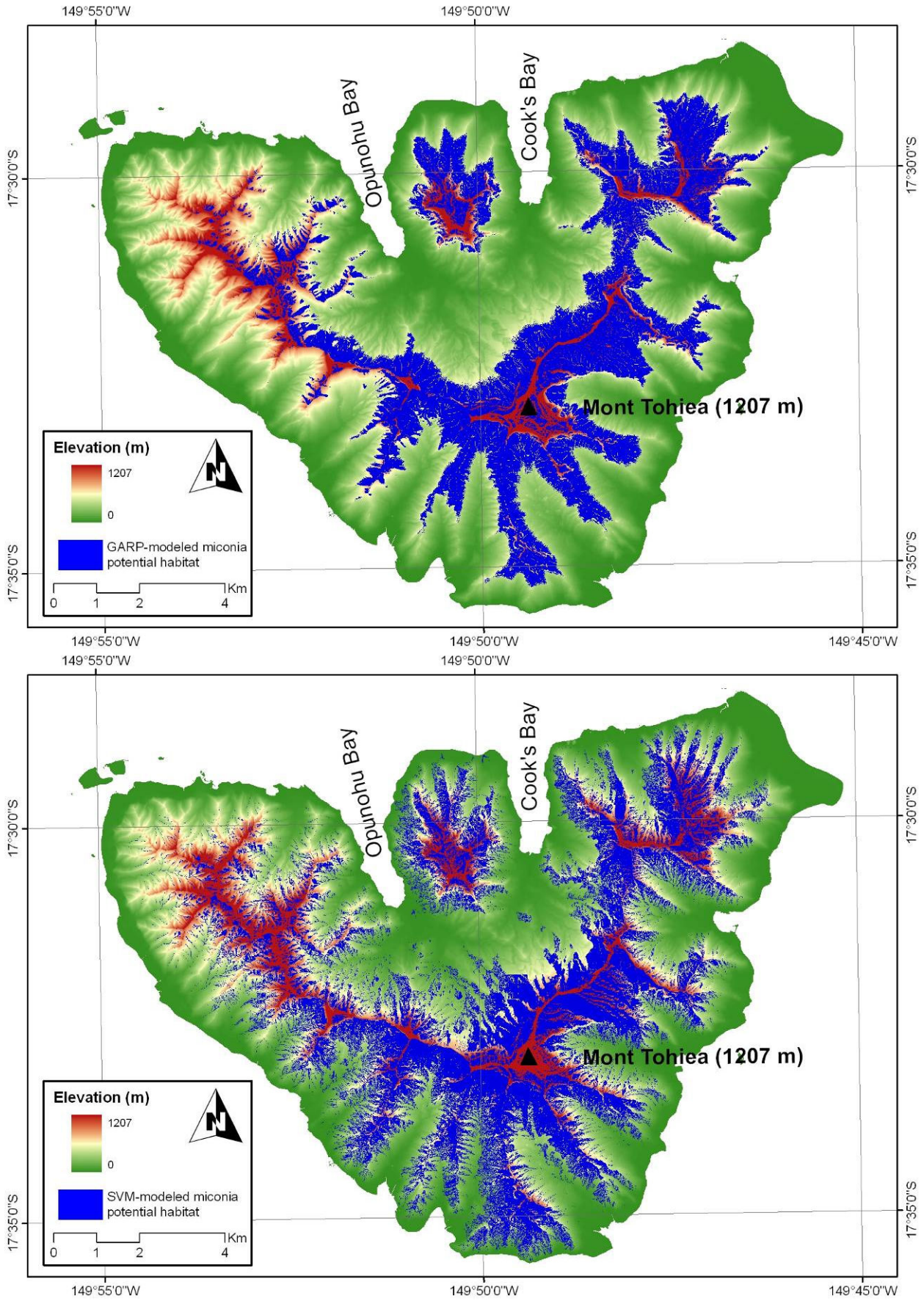
Figure 1. Comparison of miconia potential distribution modeled by GARP (above) and by a SVM regression (bellow).

REFERENCES

Anonymous, Etude sur les potentialités de l'outil télédétection à contrôler les états d'évolution de la prolifération du genre miconia à Tahiti, Délégation à l'Environnement/Station Polynésienne de Télédétection. Convention 89/1653, décembre 90, 15 p., 1990.

J. Atherton et al. (Eds), Polynesia-Micronesia biodiversity hotspot. Critical ecosystem profile. Ecosystem Partnership Fund, International-Melanesia Center for Biodiversity and Conservation. Arlington, Virginia, 2007.

C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Disc., vol. 2, no. 2, pp. 121-167, Jun. 1998.

S. K. Ebeling, E. Welk, H. Auge, H. Bruelheide, "Predicting the spread of an invasive plant: combining experiments and ecological niche model," *Ecography*, vol 31, pp. 709-719, 2008.

J. Florence, La végétation de quelques îles de Polynésie. Plates 54-55. In F. Dupon (Coord.), *Atlas de la Polynésie française.* ORSTOM (Ed), Paris, France, 1993.

G. M. Foody, A. Mathur A., "The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM," *Remote Sensing of Environment*, vol 103, pp. 179-189, 2006.

M. Fourdrigniez, J.-Y. Meyer, Liste et caractéristiques des plantes introduites naturalisées en Polynésie français. Contribution à la Biodiversité de Polynésie française no. 17, Délégation à la Recherche, Papeete, French Polynesia, 2008.

Invasive Species Specialist Group (ISSG), 2010. Available on line at www.issg.org/database/

P.E. Gessler, O.A. Chadwick, F. Chamran, L. Althouse, and K. Holmes, "Modeling soil-landscape and ecosystem properties using terrain attributes," *Soil Sci. Soc. Am. J.*, vol. 64,pp. 2046-2056, 2000.

C. Kueffer, C. C. Daehler, C. W. Torres-Santana, C. Lavergne, J.-Y. Meyer, R. Otto, L. Silva, "A global comparison of plant invasions on oceanic islands," *Perspectives in Plant Ecology, Evolution and Systematics*, vol 12, pp. 145-161, 2010.

A. M. LaRosa, M. Purrell, J. Franklin, J. Denslow, "Designing a control strategy for *Miconia calvescens* in Hawaii using spatial modeling," *9th International Conference on the Ecology and Management of Alien Plant Invasions* (poster), Perth, Australia, 2007.

S. E. Loo, R. Mac Nally, P. S. Lake, "Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges," *Ecological Applications*, vol 17, pp. 181-189, 2007.

L. L. Loope, D. Mueller-Dombois, Characteristics of invaded islands, with special references to Hawaii", In J. A. Drake et al. (Eds) *Biological invasions: a global perspective*, Chichester, John Wiley and Sons, 1989.

J.-Y. Meyer, Situation de *Miconia calvescens* à Moorea. Rapport de mission, Centre ORSTOM de Tahiti, 1993.

J.-Y. Meyer, "Threat of invasive alien plants to native flora and forest vegetation of Eastern Polynesia," *Pacific Science*, 58, 357-375, 2004.

J.-Y. Meyer, "The Miconia saga: 20 years of study and control in French Polynesia (1988-2008)," In L. L. Loope, J.-Y. Meyer, D. B. Hardesty, C. W. Smith (Eds), *Proceedings of the international Miconia conference 2009*, Keanae, Hawaii, 2010.

Millennium Ecosystem Assessment, *Ecosystems and human well being: current state and trends,* Island Press, 2005.

G. Paulay, "Biodiversity on oceanic islands its origin and extinction," *Am. Zool.*, vol 34, pp. 134-144, 1994.

A. T. Peterson, R. Scachetti-Pereira, D. A. Kluza, "Assessment of invasive potential of *Homalodisca coagulata* in Western North America and South America," *Biota Neotropica*, vol 3, pp. 1-7, 2003.

N. Roura-Pascual, A. V. Suarez, C. Gómez, P. Pons, Y. Touyama, A. L. Wild, A. T. Peterson, "Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change," *Proc. R. Soc. Lond. B*, vol 271, pp. 2527-2535, 2004.

N. Roura-Pascual, L. Brotons, A. T. Peterson, W. Thuiller, W., "Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula," *Biol. Invasions*, vol 11, pp. 1017-1031, 2008.

E. Sánchez-Flores, "GARP modelling of natural and human factors affecting the potential distribution of the invasives *Schismus arabicus* and *Brassica tournefortii* in "El Pinacate y Gran Desierto de Altar" Biosphere Reserve," *Ecological Modelling*, vol 204, pp. 457-474, 2007.

B. Schölkopf and A. Smola, *Learning with kernels*. Cambridge, MA: MIT Press, 2002.

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.

D. R. B. Stockwell, I. R. Noble, "Induction of sets of rules from animal distribution data: a robust and informative method of data analysis," *Mathematics and Computers in Simulation*, vol 32, pp. 249-254, 1991.

V. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.

B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensory data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858-3866, Dec. 2007.

Y. Xie, Z. Sha, and M. Yu, "Remote sensing imagery in vegetation mapping: a review," *J. Plant Ecology*, vol. 1, no. 1, pp. 9-23, Mar. 2008.