



**HAL**  
open science

## Support vector machines to map rare and endangered native plants in Pacific islands forests

Robin Pouteau, Ravahere Taputuarai, Jean-Yves Meyer, Benoît Stoll

► **To cite this version:**

Robin Pouteau, Ravahere Taputuarai, Jean-Yves Meyer, Benoît Stoll. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, 2012, 9, pp.37 - 46. 10.1016/j.ecoinf.2012.03.003 . hal-03791064

**HAL Id: hal-03791064**

**<https://hal.science/hal-03791064v1>**

Submitted on 13 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparing machine learning algorithms to map rare and endangered native and endemic plants in Pacific Islands forests

Pouteau Robin<sup>1,4</sup>, Ravahere Taputuarai<sup>2,4</sup>, Jean-Yves Meyer<sup>3,4</sup> & Benoît Stoll<sup>1</sup>

<sup>1</sup>*South Pacific Geosciences Laboratory, University of French Polynesia, Tahiti*

<sup>2</sup>*MaNature, Tahiti*

<sup>3</sup>*Délégation à la Recherche, Government of French Polynesia, Tahiti*

<sup>4</sup>*Moorea Biocode Project, Moorea  
French Polynesia*

## 1. Introduction

Oceanic islands are famous for their unique biota with high endemism but also their great vulnerability to anthropogenic disturbances including invasive alien species (Loope et al. 1988; Caujapé-Castells et al. 2010). As a result, a huge number of endangered species is currently found in island ecosystems (IUCN, 2001). Rare insular species may also play a key role for ecosystem functioning (Lyons & Schwartz, 2001; Lyons et al., 2005). The detailed knowledge of their ecological range and geographic distribution is thus critical for biodiversity conservation and management (Ferrier, 2002; Rushton et al., 2004).

Occurrence records are often scarce for rare species resulting in small training sample available for species distribution models (Stockwell & Peterson, 2002; Pearson et al., 2007; Wisz et al., 2008). A recent study of Williams et al. (2009) compared the ability of a range of models to predict distribution of six rare plants species (from 9 to 129 occurrences). These models included generalized linear models, artificial neural networks, the commonly used Maximum Entropy (Maxent) distribution and a Classification And Regression Tree (CART) model called Random Forests (RF) (Breiman, 2001), the latter outperforming the formers.

Nonetheless, in the field of remotely sensed data classification, a machine learning algorithm is becoming a reference: the Support Vector Machines (SVM) (Vapnik, 1998). Algorithms used in remotely sensed data classification for classifying objects reflectance are substantially the same than those used in species distribution models for classifying environmental layers (Franklin, 1995). Thus, SVM was successfully used for common species distribution modelling in few recent studies (Guo et al., 2005; Drake et al., 2006; Pouteau et al., 2011a).

SVM uses presence/absence training samples within a multidimensional feature space to fit an optimal separating hyperplane (in each dimension, vector component is an environmental layer). In this way, it tries to maximize the margin that is the distance between the closest training samples, or support vectors, and the hyperplane itself. Foody & Mathur (2006) stated that, as SVM decision boundary is only based on the few support

vectors, it is particularly able to be trained on a small meaningful sample. Nevertheless, to our knowledge, it has never been used for rare species distribution modelling.

The aim of this study is twofold: (i) to determine which model among RF and SVM is the most relevant to map rare species in a study case focusing on endangered native and endemic plants in Pacific Islands; and (ii) comparing their predicted potential habitat with their current observed range, to understand the causes of their rarity and/or endangerment.

## **2. Material and methods**

### **2.1 Target rare and endangered species**

The present study was conducted on the oceanic tropical island of Moorea (Society archipelago, French Polynesia), located at 17°30' South and 149°50' West in the South Pacific Ocean. It is a small, ca. 140 km<sup>2</sup>, high volcanic island with a rough topography and the highest summit reaching 1,207 m a.s.l. This work was part of the "Moorea Biocode Project", an international research program seeking to collect DNA sequence, distribution, morphological and ecological data of all non-microbial terrestrial and marine life in an island ecosystem (<http://www.mooreabiocode.org/>).

Three target species were selected based on their rarity and endangerment on Moorea according to their proposed IUCN status in the "Nadeaud botanical database" (Florence et al., 2007) and our personal observations from intensive field surveys during the past 4-5 years. All of the target species are found in relatively pristine forest remnants, and only the first two are legally protected in French Polynesia.

We compiled the available published and unpublished data on their location and abundance. The term "occurrence" used hereinafter refers to a 5 m x 5 m area where an isolated individual or a population of individuals is present.

#### **2.1.1 *Lepinia taitensis***

*Lepinia taitensis* (Apocynaceae) is a small tree commonly 2-5 m that grows up to 10 m in height (Fig. 1). An endemic to the islands of Tahiti and Moorea (Society archipelago), it occurs in low- to mid-elevation wet valley forests. It is listed as a "critically endangered" (CR) by IUCN Red List (IUCN, 2001). We recorded a total of 28 occurrences on Moorea.

#### **2.1.2 *Pouteria tahitensis***

*Pouteria* (syn. *Planchonella*) *tahitensis* (Sapotaceae) is a large tree, often between 10-20 m in height on Moorea (Fig. 2). It was previously described as an endemic to the Society (Florence et al., 2007), but is probably native to South Pacific islands (Swenson, U., pers. comm. 2011). It is mainly found on slopes in mid-elevation mesic to wet forests. It is considered as CR in Tahiti and Moorea (IUCN, 2001). Only 20 occurrences were recorded on Moorea.

#### **2.1.3 *Santalum insulare***

*Santalum insulare* var. *raiateense* (Santalaceae) is a shrub up to 3 m tall, endemic to the islands of Moorea and Raiatea (Society archipelago) where it occurs on low- to mid-elevation dry

and mesic ridges and slopes (Fig. 3). It is considered “near threatened” (NR) on Moorea and CR on Raiatea (IUCN, 2001). A total of 81 occurrences was recorded on the former.



Fig. 1. Habit of *Lepinia taitensis* in the understory of native lowland wet forest at 250 m elevation (R. Pouteau)



Fig. 2. A large population (20-30 individuals) of *Pouteria tahitensis* at 300 m elevation (R. Taputuarai)



Fig. 3. *Santalum insulare* on a dry and open ridge at 400 m elevation (J.-Y. Meyer)

## 2.2 Biophysical descriptors

Biophysical descriptors that can be used in species distribution models are thoroughly presented in Wilson & Gallant (2000). In the context of Pacific high volcanic islands and our target species, the descriptors that appears to us as the most contributing factors are of abiotic and biotic nature.

### 2.2.1 Abiotic

Abiotic descriptors we used include:

- elevation affecting air temperature. Considering an environmental lapse rate of  $0.0058^{\circ}\text{C}/\text{m}$  as observed in Hawaii (Baruch & Glodstein, 1999), there is a shift of  $7^{\circ}\text{C}$  between sea-level and the highest summit of Moorea (Mt Tohiea, 1,207 m). Air temperature is one of the most important factors controlling vegetation zonation and key processes such as evapotranspiration, carbon fixation and decomposition, plant productivity and mortality in mountain ecosystems (Chen et al., 1999; Nagy et al., 2003; Richardson et al., 2004);
- slope steepness (called "Slope" hereinafter) can be considered as a proxy of overland and subsurface flow velocity and runoff rate, effect of micro-topography on precipitation, geomorphology, soil water content (Wilson & Gallant, 2000), mechanical effect on plant rooting and seed dispersion;
- slope exposure (called "Aspect" hereinafter) as a proxy of solar insolation and evapotranspiration (Wilson & Gallant, 2000);
- windwardness expressing exposure to trade wind;

- a compound topographic index (*CTI*), quantifying fluid drainage by micro-topography and explaining geomorphology (Moore et al., 1993; Gessler et al., 2000), with low *CTI* values representing convex positions like mountain crests and with high *CTI* values representing concave positions like coves or hillslope bases. It is considered as a secondary physiographic descriptor since it is computed from primary physiographic descriptors (elevation and slope) (Moore et al., 1991). Indeed, *CTI* is a function of the slope angle  $\beta$  (in radians) and the specific catchment area (*As*) expressed as m<sup>2</sup> per unit width orthogonal to the flow direction (1);

$$CTI = \log(As / \tan \beta) \quad (1)$$

Abiotic descriptors were extracted from a 5 m-resolution digital elevation model (DEM).

### 2.2.2 Biotic

We used a vegetation map as the biotic descriptor. Indeed forest overstorey can affect temperature and resources (e.g. light, water, nutrients) availability by competition (Riegel et al., 1992).

Vegetation map was obtained by classifying satellite imagery. The use of both physiographic and remotely sensed data in classifications is commonly found in the literature (e.g. Strahler, 1981; Hutchinson, 1982; Linderman et al., 2004; Joshi et al., 2006). Here, satellite data is a mosaic of five 0.60 m-resolution Quickbird scenes. Very high resolution imagery such as the one from the Quickbird satellite is useful for tree species identification (Turner et al., 2003; Gillepsie et al., 2008; Xie et al., 2008). As suggested by Song et al. (2005), spectral and textural information was considered separately.

Spectral information was the native blue (430-545 nm), green (466-620 nm), red (590-710 nm) and near infrared (715-918 nm) channels.

Texture is generally referred to as the detailed spatial pattern of variability of the image average tone. Here, textural information includes eight gray-level co-occurrence matrix (GLCM) metrics introduced by Haralick (1973), namely mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation. This kind of texture metrics is the most used (Gong et al. 1992; Franklin & Peddle, 1989; Marceau et al., 1990; Aujusteijn et al., 1995; Franklin et al., 2000; Nyongui et al., 2002; Podest & Saatchi, 2002) and output good results (Paulhac, 2009). Franklin et al. (1996) and Chen et al. (2004) showed that the accuracy of classification on texture metrics can be improved by using multiple extraction window sizes. Here, texture was extracted in windows of 3x3 pixels, 9x9 pixels and 15x15 pixels visually identified as representing intra-tree micro-texture (small branches structure), intra-tree macro-texture (large branches structure) and inter-tree texture (trunks, individuals structure) respectively.

Numerous algorithms have been proposed to classify two sources such as native bands and textural information. In the comparative studies found in the literature, "winning" algorithms include the Dempster orthogonal sum combination rule (Lee et al., 1987), artificial neural networks (Benediktsson et al., 1990; Serpico & Roli, 1995), the logarithmic opinion pool (Benediktsson & Kanellopoulos), the sequential maximum a posteriori (Michelson et al., 2000), the majority voting (Fauvel et al., 2006) and SVM (Song et al., 2005; Waske &

Benediktsson, 2007; Chu & Ge, 2010). The latter was used in the most recent studies and has never been “beaten” in the previously mentioned comparative studies. The classification scheme we used is the one introduced by Waske & Benediktsson (2007): (i) native bands and textural information are classified by means of a single SVM applied on each source separately. The output of the classifier is an image containing the distance of each vector to the decision boundary (also called “rule image”); and (ii) an additional SVM is applied on the set of rule images.

Input pixels were assigned following numerous ground truth missions between 22 July 2009 and 6 February 2011. Sampled surface represents 0.2 ha for each of the 17 classes identified on Moorea, namely a total of 3.4 ha i.e. 0.25% of the island (distribution of the sample points is shown in Fig. 4). A first half of this area was used for SVM training and the other half was put aside for classification assessment.

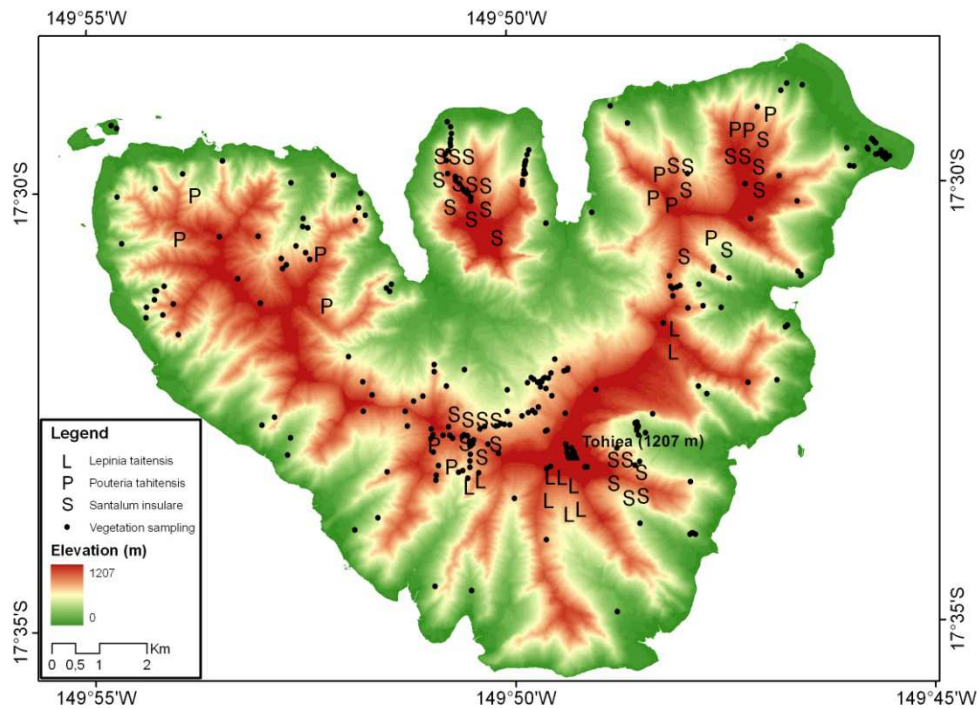


Fig. 4. Distribution of the GPS points sampled for vegetation and rare species mapping on Moorea. Locations of the three target species occurrences are deliberately shown without a precise position as they are legally protected

The resulting 0.60 m-resolution vegetation map was upscaled to a 5 m-resolution (to make easier the correspondence with abiotic descriptors) using the nearest neighbour method, more reliable than other classic methods (e.g. bilinear interpolation and cubic convolution) to resample categorical data (Baboo & Devi, 2010).

## 2.3 Machine learning algorithms

The ability of RF and SVM to integrate the aforementioned biophysical descriptors was compared. We used the implementations found in the open source machine learning software from the University of Waikato (New Zealand) called Weka (<http://www.cs.waikato.ac.nz/ml/weka/>). Rare species classifications were trained on 66% of occurrences and validated on the remaining 33%. "Absence" class is actually made of pseudo-absence pixels *i.e.* pixels randomly sampled within the unoccupied space. The same number of pixels was used for the "presence" class than for the "absence" class in order to avoid under- or over-estimation problems due to unbalanced training sets (Japkowicz & Stephen, 2002; Eitrich & Lang, 2005; Eitrich & Lang, 2006).

### 2.3.1 Random Forest

RF, introduced by Breiman (2001), is an ensemble of individual tree predictors. Each input vector is put down into each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes over all the trees in the forest. We populated our forests with the number of trees yielding the best accuracy metrics by cross-validation. RF has been recently and successfully used for species distribution modelling (Prasad et al., 2006; Cutler et al., 2007; Benito Garzon et al., 2008; Williams et al., 2009).

### 2.3.2 Support Vector Machines

Fig. 5 presents the SVM, originally introduced as a binary classifier (Vapnik, 1998) and extensively described by Burges (1998), Schölkopf & Smola (2002) and Hsu et al. (2009). Subsequently we just give a brief description of the algorithm. SVM uses two classes (e.g. presence/absence) of training samples within a multidimensional feature space to fit an optimal separating hyperplane (in each dimension, vector component is image gray-level). In this way, SVM tries to maximize the margin that is the distance between the closest training samples, or support vectors, and the hyperplane itself.

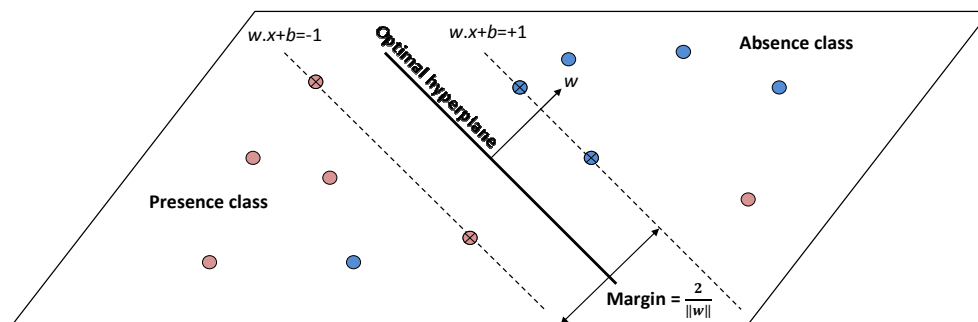


Fig. 5. Concept of a SVM. The cross circles refer to support vectors. The optimal separating hyperplane lies on them



SVM consists in projecting vectors into a high dimensional feature space by means of a kernel trick then fitting the optimal hyperplane that separates classes using an optimization function. For a generic pattern  $x$ , the corresponding estimated label  $\hat{y}$  is given by (2).

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

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 and  $a_i$  denotes the Lagrangian multipliers. Several kernels are used in the literature. According to Hsu et al. (2009) and supported by many other authors, the Gaussian Radial Basis Function (RBF) has both advantages (i) to be very successful since it works in an infinite dimensional feature space; and (ii) contrary to the other well working kernels (e.g. polynomial), RBF has a single parameter  $\gamma > 0$ . The equation is (3).

$$K(x_i, x) = \exp[-\gamma \|x_i - x\|^2] \quad (3)$$

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  $\gamma$  is found by cross-validation (Hsu et al., 2009).

Only  $a_i$  belonging to support vectors  $s_i$  have not a null value so the classification function is actually (4).

$$\hat{y} = \text{sign}[f(x)] = \text{sign}[\sum_{i=1}^{P_s} y_i a_i K(s_i, x) + b] \quad (4)$$

Wherein  $P_s$  is the number of support vectors. Thus, the decision boundary is solely based on few meaningful pixels. This is why SVM may be very appropriated for predicting distribution of species with scarce occurrence records.

### 3. Results

#### 3.1 Vegetation map

In Congalton & Green (2009), several methods for assessing classification accuracy are introduced. In our study case, we used the same two metrics as in Williams et al. (2009). They are defined as follow:

- on the one hand, AUC score refers to the area under the ROC (Receiver Operating Characteristic) curve: ROC metric is often represented by a curve corresponding to corrected assigned pixels rate according to the misclassification rate. The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% correctly assigned pixels;
- on the other hand, Cohen's Kappa expresses whether correctly assigned pixels may have been assigned by chance or not based on the classification decision rule. A value of 1 indicates perfect agreement and 0.5 indicates a pattern arising by chance.

The SVM classification of the Quickbird imagery (illustrated by Fig. 6) gives fairly good results with a Kappa of 0.842 and an AUC of 0.965. Texture is arguably the most

contributing information since the classification based on the single textural information gives a Kappa of 0.821 and an AUC of 0.955 (data not shown).

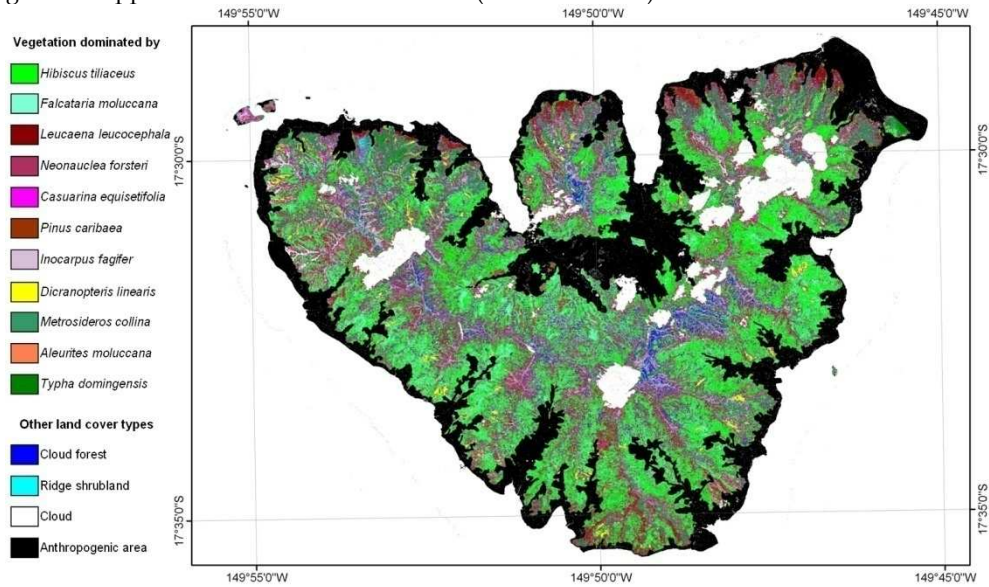


Fig. 6. Vegetation map of Moorea using a SVM classification of a Quickbird mosaic

### 3.2 Contribution of biophysical descriptors

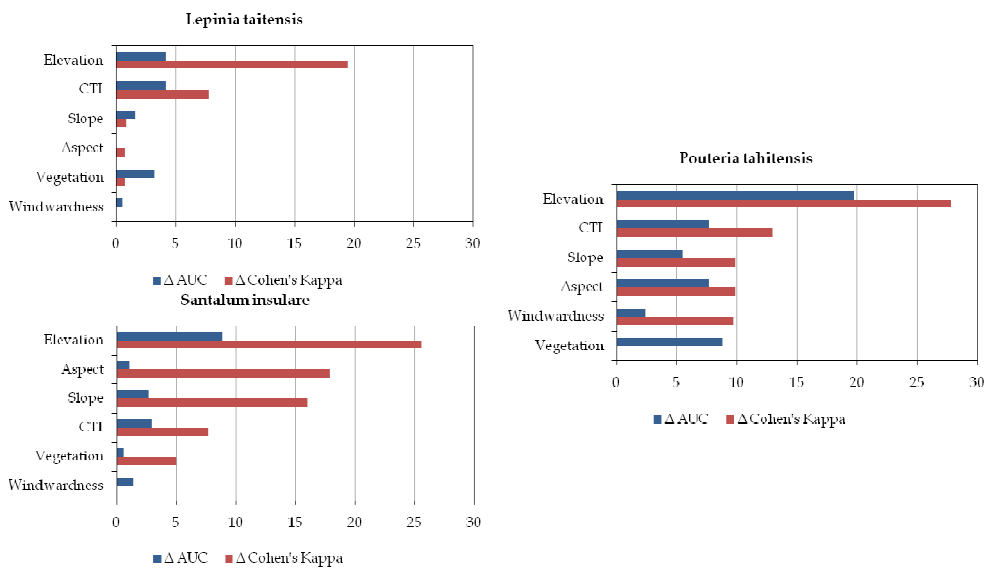


Fig. 7. Relative contribution in the SVM classification of each descriptor based on the differences of accuracy yielded with all descriptors and without the regarded descriptor

Regarding Fig. 7, descriptors relative contribution is always positive (which confirms their adaptation to the context of Pacific high volcanic islands and to our target species) but varies according to the concerned species. Although AUC and Cohen’s Kappa scores are not very correlated, the three most contributing descriptors are arguably (by order of contribution): elevation, CTI and slope. Nevertheless, overstory vegetation is conspicuously contributing to the spatial distribution of *Lepinia taitensis* which seems coherent as the species is found in the understory of dense rainforest with a probable strong dependence of incident light. Aspect is substantially influencing spatial distribution of *Santalum insulare*, which essentially occurs on dry ridges and slopes facing north.

### 3.3 Machine learning algorithms comparison

Both machine learning algorithms yield very good numerical results but, as shown by accuracy metrics presented in Table 1, SVM slightly and stably outperforms RF for the three target species.

	RF(%)	SVM (%)
(a) AUC scores		
<i>Lepinia taitensis</i>	97.4	97.9
<i>Pouteria tahitensis</i>	87.4	89.0
<i>Santalum insulare</i> var <i>raiateense</i>	97.2	97.2
(b) Cohen’s Kappa		
<i>Lepinia taitensis</i>	85.4	85.4
<i>Pouteria tahitensis</i>	70.0	78.0
<i>Santalum insulare</i> var <i>raiateense</i>	84.5	87.0

Table 1. Comparison of RF and SVM to predict occurrences for rare plant species

An “ecological niche” refers to an  $n$ -dimensional hypervolume where each dimension corresponds to the environmental range in which a species can persist (Hutchinson, 1957). SVM appears more able to fit the observed ecological niches of the three target species than RF (Fig. 8). This fact is particularly true for species with the fewest recorded occurrences: *Lepinia taitensis* (28 occurrences) and *Pouteria tahitensis* (20 occurrences). On the contrary, ecological niche of *Santalum insulare* (81 occurrences) modelled by both machine learning algorithms SVM and RF matches the observed ecological niches in a comparable way.

## 4. Discussion

### 4.1 Random Forest vs. Support Vector Machines

RF and SVM were compared on their ability to predict rare and endangered species distribution. RF was found to be optimal for predicting rare species occurrences among a wide panel of algorithms in Williams et al. (2009). SVM has never been used for predicting rare species distribution, and yet it outperforms RF in our study case. The main reason is most likely the result of the paradigm of SVM based on a small pixels sample (i.e. support vectors) (Foody & Mathur, 2006). Consequently, SVM is particularly able to be trained with

few meaningful pixels and to fit limited information. Reciprocally, SVM may not be very impacted by insignificant pixels. Now, noisy pixels are more frequent in high resolution imagery than in coarse resolution imagery thanks to information aggregation in large pixels (Hatton et al., 1997; Turner et al., 2003). SVM is thus probably more adapted than CART approaches such as RF, typically showing difficulties at a fine scale (Thuiller et al., 2003). Moreover, RF classification process occurs in a six dimensions space (one dimension per biophysical descriptor). By using the RBF kernel which works in an infinite dimensional feature space, it is easier for SVM to separate potential habitat from inappropriate habitat.

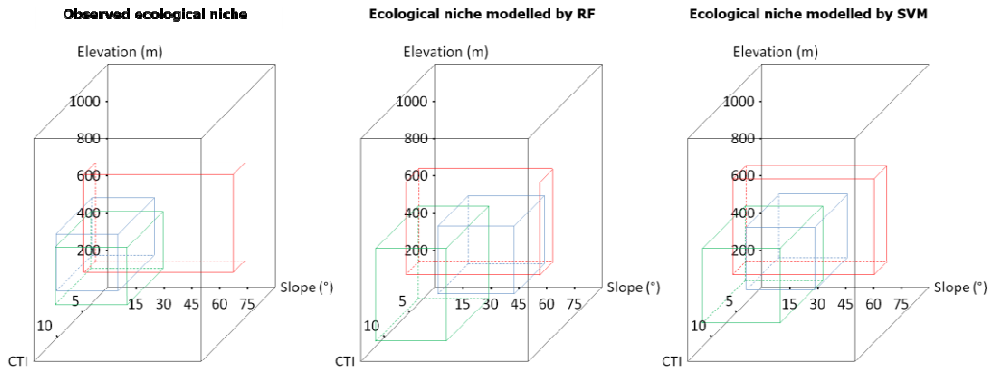


Fig. 8. Ecological niche of *Lepinia taitensis* (green), *Pouteria tahitensis* (blue) and *Santalum insulare* (red) based on the extrema of the three most contributing biophysical descriptors where the species occur.

#### 4.2 Potential habitat vs. current habitat

According to SVM, the potential distribution ranges of *Lepinia taitensis*, *Pouteria tahitensis* and *Santalum insulare* are 15 ha, 38 ha and 32 ha respectively (Fig. 9), suggesting that the target species have no very narrow and specialized habitats.

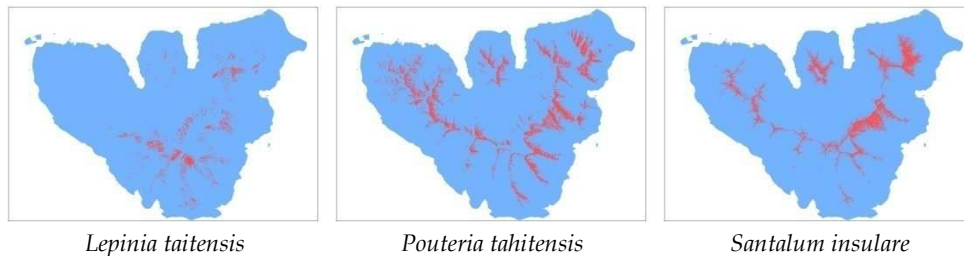


Fig. 9. Rare species distribution maps produced by SVM classification. Blue colour refers to area where the species is absent and red to its potential habitat

Current habitat based on observed occurrences are 700 m<sup>2</sup> (4.7‰ of the modelled potential habitat), 500 m<sup>2</sup> (1.3‰) and 2025 m<sup>2</sup> (6.3‰) respectively. We assume that this low density is

not related to an overestimation of the potential habitats made by the model as SVM gives fairly good analytical results as shown previously.

Moreover, the assumption that the native and endemic flora of Moorea is currently poorly known is to be excluded as the island was surveyed by naturalists and botanists in the past (Grant et al. 1974; Florence et al., 2007) and by us in an extensive way for the last 4-5 years.

We hypothesize that the discrepancy between potential and current habitats are of anthropogenic causes:

- habitat loss and fragmentation caused by deforestation, over-grazing by feral ungulates and intentional and accidental fires at low and mid-elevation since the pre-European period and in the modern times;
- invasions by alien plant species, especially in low- to mid-elevation forests (Kueffer et al., 2010). More than 180 alien species are currently naturalized on Moorea (Fourdrigniez, M., pers. comm. 2011). As an illustration, *Lepinia taitensis* and, to a lesser extent, *Pouteria tahitensis* are critically threatened by the small tree *Miconia calvescens* which overtops rainforests native flora (Meyer, 2004) and covers 25% of the island (Pouteau et al., 2011b). Other invasive plants on Moorea include the shrubs and trees *Falcataria moluccana*, *Lantana camara*, *Psidium* spp., *Tecoma stans*, *Spathodea campanulata*, *Rubus rosifolius* and *Syzygium cumini* and the vines and grasses *Melinis minutiflora*, *Merremia peltata*, *Mikania micrantha* and *Miscanthus floridulus* ;
- proliferation of introduced predatory animals. The three target species are severely depredated by rats (*Rattus* spp.) feeding on their seeds (Lhuillier et al., 2006; Meyer & Butaud, 2009);
- extirpation or extinction of endemic avian frugivores by over-hunting, predation (rats, cats, Swamp Harrier *Circus approximans*) and competition with alien birds (Red-vented Bulbul *Pycnonotus cafer*, Common Myna *Acridotheres tristis*) which lead to a lack of seed dispersal (Spotswood & Meyer, 2009).

Moreover, the three target species can be considered as *K*-strategists, characterized by a large individual size, slow life cycle and the production of a few number of large fruits (MacArthur & Wilson, 1967). Their life history traits made them more vulnerable to rapid environmental changes.

## 5. Conclusion

The first goal of this study was to compare two species ecological niche models, Random Forest (RF) with Support Vector Machines (SVM), in order to predict the distribution of rare species in natural ecosystems. We focused on three endangered native and endemic plants with small occurrence records ranging from 20 to 81 on the tropical oceanic island of Moorea (French Polynesia). Our analysis was based on six fine scale environmental layers, namely elevation, slope steepness, slope aspect, windwardness, a Compound Topographic Index (CTI) quantifying fluid drainage, and a vegetation map obtained by classifying a set of Quickbird satellite scenes. Results revealed that SVM outperforms RF especially when the number of observed occurrences is scarce. By producing more accurate maps of rare and endangered species, SVM is thus a tool of great practical value for conservation macroecologists and resource managers, and should be more considered in their future researches.

The relative high accuracy of distribution maps provided by SVM also allowed us to better understand the causes of the rarity and/or endangerment of some island native and endemic species by comparing their predicted potential habitat with their current observed range. Our data show that our three target species have a wide geographic distribution but small population size probably caused by strong anthropogenic disturbances (e.g. habitat loss and fragmentation, invasions by alien plant species and predatory animals, and possibly extirpation or extinction of natural dispersal agents) aggravated by their “outmoded” life history traits (i.e. not adapted to rapidly changing ecosystems). In the contrary, other rare and threatened species found on the island of Moorea in relatively well-preserved habitats (e.g. the subalpine vegetation or montane cloud forests) have a narrower geographic distribution but relatively large population size. We assume that the rare species found at low- and mid-elevation in the Pacific Islands are much more prone to extinction and should be of high conservation priority.

## 6. Acknowledgements

Authors are grateful to Jean-François Butaud for sharing his GPS points of the target plants, to Marie Fourdrigniez for her help during field surveys, to the Service de l’Urbanisme of the Government of French Polynesia for providing the DEM, to the Délégation à la Recherche of the Government of French Polynesia and the Moorea Biocode Project for supporting this work.

## 7. References

- Augusteijn, M. F.; Clemens, L. E. & Shaw, K. A. (1995). Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.33, pp. 616-625.
- Baboo, S. S. & Devi, R. (2010). An analysis of different resampling methods in Coimbatore, District. *Global Journal of Computer Science and Technology*, Vol.10, pp. 61-66.
- Baruch, Z. & Goldstein, G. (1999). Leaf construction cost, nutrient concentration, and net CO<sub>2</sub> assimilation of native and invasive species in Hawaii. *Oecologia*, Vol.121, pp. 183-192.
- Benediktsson, J. A.; Swain, P. H. & Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.28, pp. 540-552.
- Benediktsson, J. A. & Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.37, pp. 1367-1377.
- Benito Garzon, M.; Sanchez De Dios, R. & Sainz Ollero, H. (2008). Effects of climate change on the distribution of Iberian tree species. *Applied Vegetation Science*, Vol.11, pp. 169-178.
- Breiman, L. (2001). Random forests. *Machine Learning*, Vol.45, pp. 5-32.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, Vol.2, pp. 121-167.

- Caujapé-Castells, J.; Tye, A.; Crawford, D. J.; Santos-Guerra, A.; Sakai, A.; Beaver, K.; Lobin, W.; Florens, F. B. V.; Moura, M.; Jardim, R.; Gomes, I. & Kueffer, C. (2010). Conservation of oceanic island floras: present and future global challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, Vol.12, pp. 107-129.
- Chen, D.; Stow, D. A. & Gong, P. (2004). Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case. *International Journal of Remote Sensing*, Vol.25, pp. 2177-2192.
- Chen, J.; Saunders, S. C.; Crow, T. R.; Naiman, R. J.; Broszofske, K. D.; Mroz, G. D.; Brookshire, B. L. & Franklin, J. F. (1999). Microclimate in forest ecosystem and landscape ecology. *Bioscience*, Vol.49, pp. 288-297.
- Chu, H. T. & Ge, L. (2010). Synergistic use of multi-temporal ALOS/PALSAR with SPOT multispectral satellite imagery for land cover mapping in the Ho Chi Minh city area, Vietnam. In: *Proc. IEEE Geoscience and Remote Sensing Symposium*.
- Congalton, R. G. & Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (2nd edition). CRC Press, Boca Raton, 183 p.
- Cutler, D. R.; Edwards, T. C.; Beard, K. H.; Cutler, A. & Hess, K. T. (2007). Random forests for classification in ecology. *Ecology*, Vol.88, pp. 2783-2792.
- Drake, J. M.; Randin, C. & Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, Vol.43, pp. 424-432.
- Eitrich, T. & Lang, B. (2005). Parallel tuning of support vector machine learning parameters for large and unbalanced data sets. In: Berthold, M.R.; Glen, R.; Diederichs, K.; Kohlbacher, O. & Fischer, I., (Eds.) *Computational Life Sciences, Konstanz. Vol. 3695 of Lecture Notes in Computer Science*, Springer, pp. 253-264.
- Eitrich, T. & Lang, B. (2006). Efficient optimization of support vector machine learning parameters for unbalanced datasets. *Journal of Computational and Applied Mathematics*, Vol.196, pp. 425-436.
- Fauvel, M.; Chanussot, J. & Benediktsson, J. A. (2006). A combined support vector machines classification based on decision fusion. *Proc. IEEE Geoscience and Remote Sensing Symposium*, pp. 2494-2497.
- Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology*, Vol.51, pp. 331-363.
- Florence, J.; Chevillotte, H.; Ollier, C. & Meyer, J.-Y. (2007). Base de données botaniques Nadeaud de l'Herbier de la Polynésie française. <http://www.herbier-tahiti.pf>
- Foody, G. M. & Mathur, A. (2006). 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.
- Franklin, J. (1995). Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, Vol.19, pp. 474-499.
- Franklin, S. E. & Peddle, D. R. (1989). Spectral texture for improved class discrimination in complex terrain. *International Journal of Remote Sensing*, Vol.10, pp. 1437-1443.
- Franklin, S. E.; Wulder, M. A. & Lavigne, M. B. (1996). Automated derivation of geographic window sizes for remote sensing digital image texture analysis. *Computers and Geosciences*, Vol22, pp. 665-673.

- Franklin, S. E.; Hall, R. J.; Moskal, L. M.; Maudie, A. J. & Lavigne, M. B. (2000). Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, Vol.21, pp. 61-79.
- Gessler, P. E.; Chadwick, O. A.; Chamran, F.; Althouse, L. & Holmes, K. (2000). Modeling soil-landscape and ecosystem properties using terrain attributes. *Soil Science Society of America Journal*, Vol.64, pp. 2046-2056.
- Gillepsie, T. W.; Foody, G. M.; Rocchini, D.; Giorgi, A. P. & Saatchi, S. (2008). Measuring and modeling biodiversity from space. *Progress in Physical Geography*, Vol.32, pp. 203-221.
- Gong, P.; Marceau, D. J. & Howarth, P. J. (1992). A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment*, Vol.40, pp. 137-151.
- Grant, M. L.; Fosberg, F. R.; Smith, H. M. (1974). *Partial flora of the Society Islands: Ericaceae to Apocynaceae*. Smithsonian contributions to botany, No. 17, Smithsonian Institution Press, Washington, 85 p.
- Guo, Q.; Kelly, M. & Graham, C. H. (2005). Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, Vol.182, pp. 75-90.
- Haralick, R. M.; Shanmugam, K. & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, Vol.3, pp. 610-620.
- Hatton, T. J.; Salvucci, G. D. & Wu, H. I. (1997). Eagleson's optimality theory of an ecohydrological equilibrium: quo vadis? *Functional Ecology*, Vol.11, pp. 665-674.
- Hsu, C. W.; Chang, C. C.; Lin, C. J. (2009). A practical guide to support vector classification. Technical Note, Department of Computer Science and Information Engineering, National Taiwan Univ., Taiwan.
- Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, Vol.22, pp. 415-427.
- Hutchinson, C. F. (1982). Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing*, Vol.48, pp. 123-130.
- IUCN (International Union for Conservation of Nature) (2001). *IUCN red list and criteria: IUCN species survival commission*. Version 3.1. IUCN, Gland, Switzerland and Cambridge, UK, pp. 2-30.
- Japkowicz, N. & Stephen, S. (2002). The class imbalance problem: a systematic study. *Intelligent Data Analysis*, Vol.6, pp. 429-450.
- Joshi, C.; De Leeuw, J.; van Andel, J.; Skidmore, K. A.; Lekhak, H. D.; van Duren, I. C. & Norbu, N. (2006). Indirect remote sensing of a cryptic forest understorey invasive species. *Forest Ecology and Management*, Vol.225, pp. 245-256.
- Kueffer, C.; Daehler, C. C.; Torres-Santana, C. W.; Lavergne, C.; Meyer, J.-Y.; Otto, R. & Siolva, L. (2010). A global comparison of plant invasions on oceanic islands. *Perspectives in Plant Ecology, Evolution and Systematics*, Vol.12, pp. 145-161.
- Lee, T.; Richards, J. A. & Swain, P. H. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.25, pp. 283-293.



- Lhuillier, E.; Butaud, J.-F. & Bouvet, J.-M. (2006). Extensive clonality and strong differentiation in the insular Pacific tree *Santalum insulare*: implications for its conservation. *Annals of Botany*, Vol.98, pp. 1061-1072.
- Linderman, M. A.; Liu, J.; Qi, J.; An, L.; Ouyang, Z.; Yang, J. & Tan, T. (2004). Using artificial neural networks to map the spatial distribution of understory bamboo from remotely sensed data. *International Journal of Remote Sensing*, Vol.25, pp. 1685-1700.
- Loope, L. L.; Hamann, O. & Stone, C. P. (1988). Comparative conservation biology of oceanic archipelagoes. *BioScience*, Vol.38, pp. 272-282.
- Lyons, K. G. & Schwartz, M. W. (2001). Rare species loss alters ecosystem function - invasion resistance. *Ecology Letters*, Vol.4, pp. 358-365.
- Lyons, K. G.; Brigham, C. A.; Traut, B. H. & Schwartz, M. W. (2005). Rare species and ecosystem functioning. *Conservation Biology*, Vol.19, pp. 1019-1024.
- MacArthur, R. & Wilson, E. O. (1967). *The theory of island biogeography*, Princeton University Press, Princeton, 203 p.
- Marceau, D. J.; Howarth, P. J.; Dubois, J. M. & Gratton, D. J. (1990). Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.28, pp. 513-519.
- Meyer, J.-Y. (2004). Threat of invasive alien plants to native flora and forest vegetation of Eastern Polynesia. *Pacific Science*, Vol.58, pp. 357-375.
- Meyer, J.-Y. & Butaud, J.-F. (2009). The impacts of rats on the endangered native flora of French Polynesia (Pacific Islands): driver of plant extinction or coup de grace species? *Biological Invasions*, Vol.11, pp. 1569-1585.
- Michelson, D. B.; Liljeberg, B. M. & Pilesjo, P. (2000). Comparison of algorithms for classifying Swedish landcover using Landsat TM and ERS-1 SAR data. *Remote Sensing of Environment*, Vol.71, pp. 1-15.
- Moore, I. D., Grayson, R. B., Ladson, A. R. (1991). Digital terrain modelling: a review of hydrological, geomorphological and biological applications. *Hydrological Processes*, Vol.5, pp. 3-30.
- Nagy, L.; Grabherr, G.; Körner, C. & Thompson, D. B. A. (Eds.) (2003). *Alpine Biodiversity in Europe*. Springer Verlag, Berlin.
- Nyoungui, A.; Tonye, E. & Akono, A. (2002). Evaluation of speckle filtering and texture analysis methods for land cover classification from SAR images. *International Journal of Remote Sensing*, Vol.23, pp. 1895-1925.
- Paulhac, L. (2009). Outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques. PhD thesis of the Univ. Tours François Rabelais (France).
- Pearson, R. G.; Raxworthy, C. J.; Nakamura, M. & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a case using cryptic geckos in Madagascar. *Journal of Biogeography*, Vol.34, pp. 102-117.
- Podest, E. & Saatchi, S. (2002). Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing*, Vol.23, pp. 1487-1506.
- Pouteau, R.; Meyer, J.-Y. & Stoll, B. (2011). A SVM-based model for predicting distribution of the invasive tree *Miconia calvescens* in tropical rainforests. *Ecological Modelling*, Vol.222, pp. 2631-2641.

- Pouteau, R.; Meyer, J.-Y.; Taputuarai, R. & Stoll, B. (2011). A comparison between GARP model and SVM regression to predict species potential distribution: mapping the invasive *Miconia calvescens* on Moorea, French Polynesia. *Proc. International Symposium for Remote Sensing of Environment*.
- Prasad, A. M.; Iverson, L. R. & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, Vol.9, pp. 181-199.
- Richardson, A. D. (2004). Foliar chemistry of balsam fir and red spruce in relation to elevation and the canopy light gradient in the mountains of the northeastern United States. *Plant and Soil*, Vol.260, pp. 291-299.
- Riegel, G. M.; Miller, R. F. & Krueger, W. C. (1992). Competition for resources between understory vegetation and overstory *Pinus ponderosa* in Northeastern Oregon. *Ecological Applications*, Vol.2, pp. 71-85.
- Rushton, S. P.; Ormerod, S. J. & Kerby, G. (2004). New paradigm for modeling species distribution? *Journal of Applied Ecology*, Vol.41, pp. 193-200.
- Schölkopf, B.; Smola, A. (2002). *Learning with kernels*. MIT Press, Cambridge.
- Serpico, S. B. & Roli, F. (1995). Classification of multisensor remote-sensing images by structured neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.33, pp. 562-578.
- Song, X.; Fan, G. & Rao, M. (2005). Automatic CRP mapping using nonparametric machine learning approaches. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, pp. 888-897.
- Spotswood, E. N. & Meyer, J.-Y. (2009). Interactions between plants and avian frugivores in the Society Archipelago, French Polynesia. *Proc. Pacific Science Inter-Congress*.
- Stockwell, D. R. B. & Peterson, A. T. (2002). Effect of sample size on accuracy of species distribution models. *Ecological Modelling*, Vol.143, pp. 1-13.
- Strahler, A. H. (1981). Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data. *International Journal of Remote Sensing*, Vol.2, pp. 15-41.
- Thuiller, W.; Araujo, M. B. & Lavorel, S. (2003). Generalized models vs. Classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, Vol.14, pp. 669-680.
- Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E. & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, Vol.18, pp. 306-314.
- Vapnik, V. (1998). Statistical learning theory. In: *Support vector machines for pattern recognition*. New York: John Wiley & Sons.
- Waske, B. & Benediktsson, J. A. (2007). Fusion of support vector machines for classification of multisensor data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.45, pp. 3858-3866.
- Williams, J. N.; Seo, C.; Thornes, J.; Nelson, J. K.; Erwin, S.; O'Brien, J. M. & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, Vol.15, pp. 565-576.
- Wilson, J. P. & Gallant, J. C. (2000). *Terrain analysis: principles and applications*. New York: John Wiley & Sons.

- Wisz, M. S.; Hijmans, R. J.; Li, J.; Peterson, A. T.; Graham, C. H. & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, Vol.14, pp. 763-773.
- Xie, Y.; Sha, Z. & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Plant Ecology*, Vol.1, pp. 9-23.