Escobar-Flores JG, Lopez-Sanchez CA, Sandoval S, Marquez-Linares MA, Wehenkel C. 2018. Predicting *Pinus monophylla* forest cover in the Baja California Desert by remote sensing. PeerJ 6:e4603
https://doi.org/10.7717/peerj.4603

Predicting *Pinus monophylla* Forest in the Baja California Desert by Remote Sensing

Jonathan. G. Escobar-Flores [1], Carlos A. López-Sánchez [2], Sarahi Sandoval [3], Marco A. Márquez-Linares [1], Christian Wehenkel [2]

[1] Instituto Politécnico Nacional. Centro Interdisciplinario De Investigación para el Desarrollo Integral Regional, Unidad Durango., Durango, México

[2] Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Durango, México

[3] CONACYT - Instituto Politécnico Nacional. CIIDIR. Unidad Durango. Durango, México

Corresponding author:

Christian Wehenkel [2]

Km 5.5 Carretera Mazatlán, Durango, 34120 Durango, México

Email address: wehenkel@ujed.mx

ABSTRACT

Background. The Californian single-leaf pinyon (*Pinus monophylla* var. *californiarum*), a subspecies of the single-leaf pinyon (the world's only 1-needled pine), inhabits semi-arid zones of the Mojave Desert (southern Nevada and southeastern California, US) and also of northern Baja California (Mexico). This subspecies is distributed as a relict in the geographically isolated arid Sierra La Asamblea at elevations of between 1,010 and 1,631 m, with mean annual precipitation levels of between 184 and 288 mm. The aim of this research was i) to estimate the distribution of *P. monophylla* var. *californiarum* in Sierra La Asamblea, Baja California (Mexico) by using Sentinel-2 images, and ii) to test and describe the relationship between the distribution of *P. monophylla* and five topographic and 18 climate variables. We hypothesized that i) Sentinel-2 images can be used to predict the *P. monophylla* distribution in the study site due to higher resolution (x3) and increased number of bands (x2) relative to Landsat-8 which is

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

public free of charge and have been proved useful for estimating forest cover, and ii) the topographical variables aspect, ruggedness and slope are particularly important because they represent important microhabitat factors that can determine where conifers can become established and persist. **Methods.** An atmospherically corrected a 12-bit Sentinel-2A MSI image with ten spectral bands in the visible, near infrared, and short-wave infrared light region was used in combination with the normalized differential vegetation index (NDVI). Supervised classification of this image was carried out using a backpropagation-type artificial neural network algorithm (BPNN). Stepwise multivariate binominal logistical regression and Random Forest classification including cross valuation (10-fold) were used to model the associations between presence/absence of *P. monophylla* and the five topographical and 18 climate variables. **Results.** We estimated, using supervised classification of Sentinel-2 satellite images, that *P. monophylla* covers 5,395 ±23.29 hectares in the isolated Sierra La Asamblea. The NDVI was one of the variables that contributed to the prediction and clearly separated the forest cover (NDVI > 0.35) from the other vegetation cover (NDVI < 0.20). The ruggedness was the most influential environmental predictor variable and indicated that the probability of *P. monophylla* occurrence was higher than 50% when the degree of ruggedness was greater than 17.5 m. When average temperature in the warmest month increased from 23.5 to 25.2 °C, the probability of occurrence of *P. monophylla* decreased. **Discussion.** The classification accuracy was similar to that reported in other studies using Sentinel-2A MSI images. Ruggedness is known to generate microclimates and provides shade that decreases evapotranspiration from pines in desert environments. Identification of *P. monophylla* in the Sierra La Asamblea as the most southern populations represents an opportunity for research on climatic tolerance and community responses to climate variability and change.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

INTRODUCTION

The Californian single-leaf pinyon (*Pinus monophylla* var. *californiarum*), a subspecies of the single-leaf pinyon (the world's only 1-needled pine), inhabits semi-arid zones of the Mojave Desert (southern Nevada and southeastern California, US) and also of northern Baja California (BC) (Mexico). It is cold-tolerant, drought resistant and is mainly differentiated from the typical subspecies *Pinus monophylla* var. *monophylla* by a larger number of leaf resin canals and longer fascicle-sheath scales (Bailey, 1987). This subspecies was first reported in BC in 1767 (Bullock et al., 2006). The southernmost record of *P. monophylla* var. *californiarum* in America was previously in BC, 26-30 miles north of Punta Prieta, at an elevation of 1,280 m (longitude -114°.155; latitude 29°.070, catalogue number ASU 0000235), and the type specimen is held in the Arizona State University Vascular Plant Herbarium.

This subspecies is distributed as a relict in the geographically isolated Sierra La Asamblea, at a distance of 196 km from the Southern end of the Sierra San Pedro Martir and at elevations of between 1,010 and 1,631 m (Moran, 1983, Table 2), with mean annual precipitation levels of between 184 and 288 mm (Roberts & Ezcurra, 2012, Table 2). The Californian single-leaf pinyon grows together with up to about 86 endemic plant species, although the number of species decreases from north to south (Bullock et al., 2008).

Adaptation of *P. monophylla* var. *californiarum* to arid ecosystems enables the species to survive annual precipitation levels below 150 mm. In fact, seeds of this variety display a high survival rate under shrubs such as *Quercus spp.* and *Arctostaphylus spp.,* a strategy that enables the pines to widen their distribution, as has occurred in the great basin in California (Callaway et al., 1996; Chambers, 2001) and for them to occupy desert zones such as Sierra de la Asamblea. Despite the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

importance of this relict pine species, its existence is not considered in most forest inventories in Mexico (CONABIO, 2017).

Remote sensing with Landsat images has proved useful for estimating forest cover; the Landsat-8 satellite has sensors (7 bands) that can be used to analyze vegetation at spatial resolution of 30 m (Madonsela et al., 2017). However, the European Space Agency's Copernicus program has made Sentinel-2 satellite images available to the public free of charge. The spatial resolution (10 m per pixel) is three times greater than that of Landsat images, thus increasing their potential for predicting and differentiating types of vegetation cover (Drush et al., 2012; Borras et al., 2017). The Sentinel-2 has 13 bands, of which 10 provide high-quality radiometric images of spatial resolution 10-20 m in the visible and infrared regions of the electromagnetic spectrum. These images are therefore ideal for land classification (ESA, 2017).

The aim of this research was i) to estimate the distribution of *Pinus monophylla* var. *californiarum* in the Sierra La Asamblea, Baja California (Mexico) by using Sentinel-2 images, and ii) to test and describe the relationship between this distribution of *P. monophylla* and five topographic and 18 climate variables. We hypothesized that i) the Sentinel-2 images can be used to accurately predict the *P. monophylla* distribution in the study site due to finer resolution (x3) and increased number of bands (x2) relative to Landsat-8, and ii) the topographical variables aspect, ruggedness and slope are particularly influential because they represent important microhabitat factors that can affect where conifers can become established and persist (Marston, 2010).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

MATERIALS AND METHODS

Study area

Sierra La Asamblea is located in Baja California's central desert (-114° 9' W 29° 19´ N, elevation range 280-1,662 m, Fig. 1). The climate is arid, with maximum temperatures of 40° C in the summer (Garcia, 1998). The Sierra is steeper on the western slopes, with an average incline of 35º, and with numerous canyons with occasional springs and oases. Valleys and plateaus are common in the proximity of the Gulf of California. Granite rocks occur south of the Sierra and meta-sedimentary rocks along the north and southeast of the slopes. The predominant types of vegetation are xerophilous scrub, which is distributed at elevations ranging from 200 to 1,000 m. Chaparral begins at an altitude of 800 m, and representative specimens of *Adenostoma fasciculatum, Ambrosia ambrosioides, Dalea bicolor orcuttiana Quercus tuberculata, Juniperus california* and *Pinus monophylla* are also present at elevations higher than 1,000 m. Populations of the endemic palm tree *Brahea armata* also occur in the lower parts of the canyons with superficial water flow and through the rocky granite slopes (Bullock et al., 2006).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

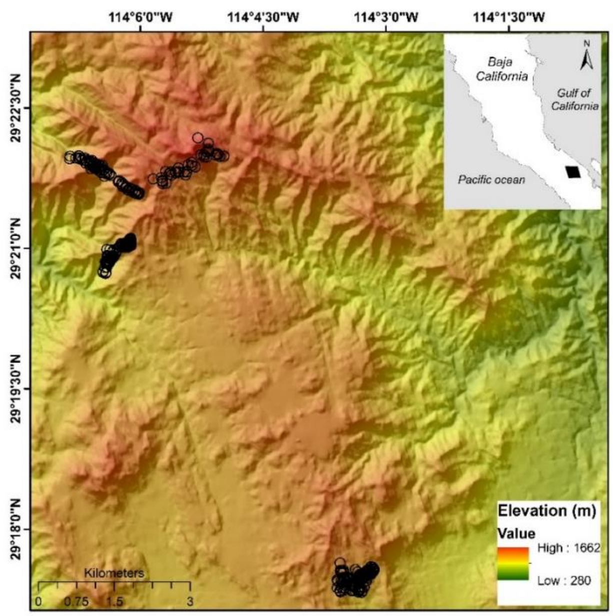


Figure 1. Map of Sierra La Asamblea. The black circles indicate georeferenced sites occupied by *Pinus monophylla*.

Datasets

Sentinel-2

The Sentinel-2A multispectral instrument (MSI) L1C dataset, acquired on 11 October 2016, in the trajectory of coordinates latitude 29º.814, longitude 114º.93, was downloaded from the US. Geological Survey (USGS) Global Visualizaton Viewer at http://glovis.usgs.gov/. The 12-bit Sentinel-2A MSI image has 13 spectral bands in the visible, NIR, and SWIR wavelength regions with spatial resolutions of 10-60 m. However, the band one used for studies of coastal aerosol and bands nine and ten, applied for respectively water vapour correction and cirrus detection, were not used in this study (ESA, 2017). Hence, the data preparation involved four bands at 10 m and the resampling of the six S2 bands acquired at 20 m to obtain a layer stack of 10 spectral bands at 10 m (Table 1) using the ESA´s Sentinel-toolbox ESA Sentinel Application Platform (SNAP) and then converted to ENVI format.

Because atmospherically improved images are crucial for assessing spectral indices with spatial reliability and product comparison, Level-1C data were converted to Level-2A (Bottom of Atmosphere -BOA- reflectance) taking into account the effects of aerosols and water vapour on reflectance (Radoux et al., 2016). These corrections were made using the Sen2Cor tool (Telespazio VEGA Deutschland GmbH, 2016) for Sentinel-2 images.

Table 1. Sentinel-2 spectral bands used to predict the *Pinus monophylla* forest

Band	Central wavelength (µm)	Resolution (m)
Band 2–Blue	0.490	10
Band 3 –Green	0.560	10
Band 4 – Red	0.665	10
Band 5- Vegetation red edge	0.705	20
Band 6– Vegetation red edge	0.740	20
Band 7– Vegetation red edge	0.783	20
Band 8- NIR	0.842	10
Band 8A– Vegetation red edge	0.865	20
Band 9 – Water vapour	0.945	60
Band 11 –SWIR	1.610	20
Band 12 –SWIR	2.190	20

The following equation was used to calculate the normalized difference vegetation index (NDVI): NDVI = (NIR – R) / (NIR + R), where NIR is the near infrared light (band) reflected by the vegetation, and R is the visible red light reflected by the vegetation (Rouse et al., 1974). The NDVI is useful for discriminating the layers of temperate forest from scrub and chaparral. Areas occupied by large amounts of unstressed green vegetation will have values much larger than 0 and areas with no vegetation will have values close to 0 and, in some cases, negative values (Pettorelli, 2013). The NDVI image was combined with the previously described multi spectral bands.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

Environmental variables

Tree species distribution is generally modulated by hydroclimate and topographical variables (Elliot et al., 2005; Decastilho et al., 2006), which can be determined from digital terrain models (DTM) (Osem et al., 2005; Spasojevic et al., 2016). A DTM was obtained by using tools available from the Instituto Nacional de Estadistica y Geografía (http:www.inegi.org.mx/geo/contenidos/datosrelieve) with a spatial resolution of 15 m. The DTM was processed with the QGIS (QGIS Development Team, 2016), using *Terrain analysis* tools, elevation, slope and aspect (Table 2).

The ruggedness was estimated using two indexes: i) the terrain ruggedness index (TRI) of Riley et al. (1999) and ii) a vector ruggedness measure (VRM), both implemented in QGIS (QGIS Development Team, 2016). The TRI computes the values for each grid cell of a DEM. This calculates the sum change in elevation between a grid cell and its eight-neighbor grid cell. VRM incorporates the heterogeneity of both slope and aspect. This measure of ruggedness uses 3-dimensional dispersion of vectors normal to planar facets on landscape. This index lacks units and ranges from 0 (indicating a totally flat area) to 1 (indicating maximum ruggedness) (Sappington et al., 2007).

On the other hand, 18 climate variables with a 30-arc second resolution (approximate 800 meters) (Table 2) were obtained from a national database managed by the University of Idaho (http://charcoal.cnre.vt.edu/climate) and which requires point coordinates (latitude, longitude and elevation) as the main inputs (Rehfeldt, 2006; Rehfeldt et al., 2006). These variables are frequently used to study the potential effects of global warming on forests and plants in Western North America and Mexico (Sáenz-Romero et al., 2010; Silva-Flores et al., 2014).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

Table 2. Topographical and climatic variables considered in the study

Variable	Abbreviation	Units	Mean	SD	Max	Min
Ruggedness	IRT	m	20.33	6.66	35.90	4.69
Ruggedness VRM	VRM	NA	0.005	0.007	0.13	0
Slope	S	°	28.38	8.92	48.34	3.42
Aspect *	A	°	190.51	68.72	350.44	20.55
Elevation *	E	m	1302.41	124.96	1631	1010
Mean annual temperature *	MAT	°C	16.57	0.38	17.4	15.5
Mean annual precipitation *	MAP	mm	229.56	19.95	288	184
Growing season precipitation, April-September *	GSP	mm	79.08	9.60	108	57
Mean temperature in the coldest month *	MTCM	°C	10.85	0.37	11.7	9.8
Minimum temperature in the coldest month *	MMIN	°C	3.42	0.41	4.3	2.3
Mean temperature in the warmest month	MTWM	°C	24.52	0.31	25.2	23.5
Maximum temperature in the warmest month	MMAX	°C	34.10	0.31	34.7	33.1
Julian date of the last freezing data of spring *	SDAY	Days	82.57	7.86	106	60
Julian date of the first freezing data of autumn *	FDAY	Days	331.28	2.62	339	324
Length of the frost-free period *	FFP	Days	259.22	8.36	285	240
Degree days > 5°C *	DD5	Days	4245.26	137.52	4550	3852
Degree days > 5°C accumulating within the frost-free period *	GSDD5	Days	3491.82	164.76	3944	2995
Julian date when the sum degree days > 5°C reaches 100 *	D100	Days	17.07	1.10	20	15
Degree days < 0 °C *	DD0	Days	0	0	0	0
Minimum degree days < 0 °C *	MMINDD0	Days	8.07	20.29	145	45
Spring precipitation	Sprp	mm	7.54	0.71	10	6
Summer precipitation *	Smrp	mm	43.74	6.29	62	29
Winter precipitation *	Winp	mm	110.93	7.93	133	93

* Variables for which no significant difference between the medians was obtained after Bonferroni correction ($\alpha = 0.0005$) were excluded from further analysis.

Pixel-based classification

Classification method

Pixel-based classification was carried out in order to identify four different types of land cover in the study area (*P. monophylla*, scrub, chaparral and no apparent vegetation), using a supervised classification approach with a backpropagation-type artificial neural network (BPNN) (SNAP, 2017). BPNN is widely used because of its structural simplicity and robustness in modelling non-linear relationships. In this study, the BPNN comprises a set of three layers (raster): an input layer, a hidden layer and an output layer (Richards, 1993). Each layer consists of a series of parallel processing elements (neurons or nodes). Each node in a layer is linked to all nodes in the next layer (Guo et al., 2013).

The first step in BPNN supervised classification is to enter the input layer, which in this study corresponded to the values of the pixels of ten Sentinel-2 bands and of the NDVI image. Weights were then assigned to the BPNN to produce analytical data from the input values. These data were contrasted with the category to which each training pixel belongs, corresponding to Georeferenced sites (Datum WGS-84, 11N) obtained in the field in October 2014 and October 2015.

A stratified random sampling method (Olofsson et al., 2013) was used to generate the reference data in the software QGIS (QGIS Development Team 2016). A total of 4017 random points were sampled, with at least 400 points for each class (Goodchild et al., 1994). The following classes were considered: i) *P. monophylla*, 502 sites, ii) scrub, 563 sites, iii) chaparral, 419 sites, and iv) no apparent vegetation, 419 sites. Class discrimination processes occurred in the hidden layer and the synapses between the layers were identified by an activation function. We used a logistic

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

function and training rate of 0.20, previously used for land cover classification (Hepner et al., 1990; Richards, 1993; Braspenning & Thuijisman, 1995). Learning occurs by adjusting the weights in the node to minimize the difference between the output node activation, and BPNN then calculates the error at each iteration with root square error (RMS). The output layer comprised four neurons representing the four target classes of land cover (*P. monophylla*, scrub, chaparral and no apparent vegetation).

Validation

The BPNN classification was cross-validated (10-fold) using a confusion matrix, which is a table that compares the reference data and the classification results. The confusion matrix was also used to determine the overall accuracy (the proportion of the area mapped correctly), user accuracy (proportion of the area mapped as a particular category that is actually that category) and producer accuracy (proportion of the area that is a particular category on the ground that is also mapped as that category) (Congalton, 1991). We calculated the uncertainty of the classification through estimated error matrix with 95% confidence intervals. We then generated a map from the results of the probability of class assignment. Finally, we estimated the area of *P. monophylla* and calculated the standard error, error-adjusted and 95% confidence intervals proposed by Olofsson et al. (2013).

The accuracy of classification was also calculated using the Kappa (*K*) coefficient. The *K* coefficient is often used as an overall measure of accuracy. This coefficient takes values of between 0 and 1, where values close to one indicate a high degree of agreement between classes and observations, and a value of 0 suggests that the observed agreement is random (Abraira, 2001). However, the use of *K* is controversial because i) *K* would underestimate the probability

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

that a randomly selected pixel is correctly classified, ii) K is highly correlated with overall accuracy so reporting Kappa is redundant for overall accuracy (Olofsson et al., 2014).

Relationship between presence of *P. monophylla* and environmental variables

To model and test the association between presence/absence of *P. monophylla* in the study area and topographical or climate variables (Table 2) a Kruskal-Wallis test was used to determine the difference in the median values in relation to presence and absence of *P. monophylla*. All variables for which no significant difference between the medians was obtained after Bonferroni correction (α = 0.0005) were excluded from further analysis. The collinearity between the variables with a significant difference between the medians of presence and absence was measured using the Spearman correlation coefficient (r_s). When the r_s value for the difference between two variables was larger than 0.7, only the variable with the lowest p value in the Kruskal-Wallis test was used in the multivariate models (as reported by Salas et al., 2017 and Shirk et al., 2018). Finally, stepwise multivariate binominal logistical regression and Random Forest classification including cross valuation (10-fold) were used to model the associations between presence/absence of *P. monophylla* and the most important topographical and climate variables (Shirk et al., 2018).

Regression and classification including cross-validations were carried out using the trainControl, train, glm (family = "binomial") and rf functions, as well as the “randomForest” and “caret” packages (Venables and Ripley, 2002) in R (version 3.3.2) (Development Core Team, 2017). The goodness-of-fit of the logistical regression model was evaluated using Akaike information criterion (AIC), root-mean-square error (RMSE) and residual deviance. Validation of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

RandomForest model was performed using under the curve (AUC; Fawcett, 2006), True Skill Statistic (TSS; Allouche et al., 2006), Kappa (Abraira, 2001), specificity and sensitivity.

RESULTS

Pixel-based classification

We estimated the area of *P. monophylla* cover of 5,395 ± 23.29 hectares in the in Sierra de la Asamblea, Baja California, Mexico. The supervised classification with BPNN yielded predictions with an overall accuracy of identification of 89.78%. This level of accuracy was obtained in the 32 interactions with 0.04 RMS training. The proportion of omission errors in the pine class was only 12.42%, *i.e.* 87.58% of the pixels were correctly classified. The chaparral class had the larger proportion of omission errors (27.65%) (Table 3, Fig. 2; Fig. 3, Fig. 4). The value of NDVI in the *P. monophylla* forest fluctuated between 0.30 and 0.41, and in chaparral between 0.24 and 0.28. The lowest values of NDVI corresponded to scrub vegetation, with values between 0.10 and 0.15.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

Table 3. Results of the classification monitored by BPNN. The overall accuracy of classification was 89.78%.

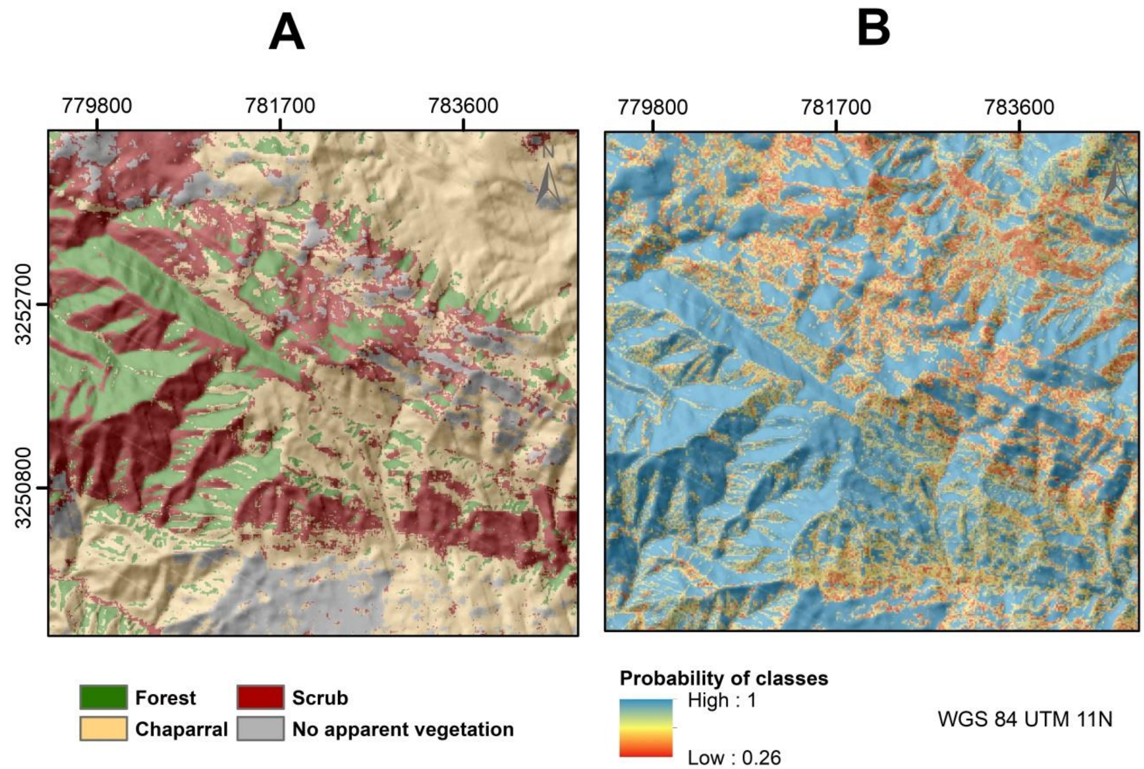
	Reference data (Known Cover Types) *					Accuracy (%)	
Classification data	**P**	**S**	**C**	**WV**	**Total**	**Producer´s**	**User´s**
P	522	0	14	0	536	87.58	97.39
S	24	619	119	2	764	100	81.02
C	50	0	348	7	405	72.35	85.93
WV	0	0	20	418	438	97.85	100
Total	596	619	481	418	2,143		

* P = piñon pine; S = shrub; C = chaparral; WV= without vegetation

Table 4. Estimated error matrix based on Table 3 with cell entries expressed as the estimated proportion of area. Accuracy measures are presented with a 95% confidence interval. Map categories are the rows while the reference categories are the columns.

Classification data	P	S	C	WV	User´s	Producer´s	Overall
P	0.244	0.000	0.007	0.000	0.96±0.97	0.82±0.91	0.89±0.01
S	0.011	0.290	0.056	0.001	------	------	
C	0.023	0.000	0.162	0.003	0.84±0.87	0.61±0.76	
WV	0.000	0.000	0.009	0.196	0.94±0.96	0.95±0.98	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018



253
254 **Figure 2**. (A) Estimated land cover classes using BPNN classification in the Sierra La Asambla.
255 (B) Probability map of class assignment. Sentinel-2 has a greater capacity for the correct
256 discrimination of the coverage corresponding to the *Pinus monophylla.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

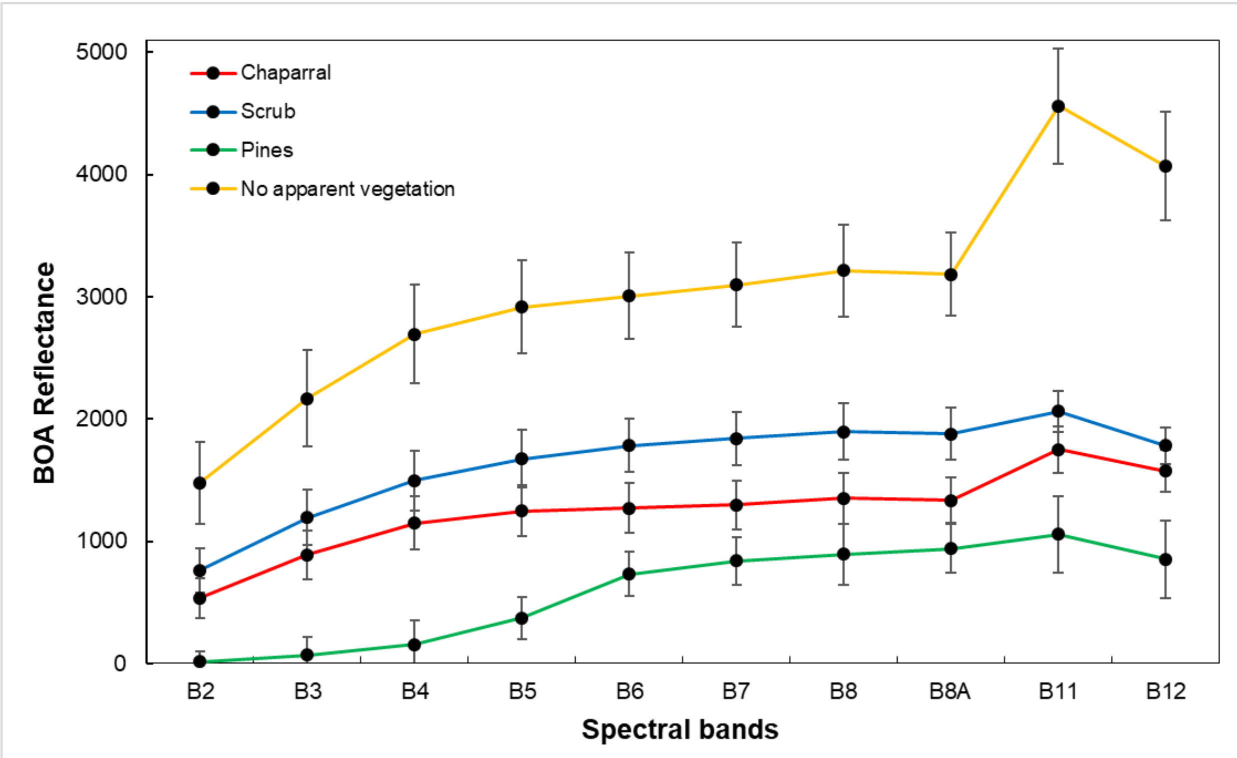


Figure 3. Spectral signatures of cover vegetation in the Sierra La Asamblea, Baja California.

Relationship between presence of *P. monophylla* and environmental variables

The Kruskal-Wallis test indicated that the median values for ruggedness TRI ($p < 2.1x10^{-16}$), slope ($p < 2.2x10^{-16}$), ruggedness VRM ($p = 4.9x10^{-9}$), MTWM ($p = 0.000014$), MMAX ($p = 0.000048$) and SPRP ($p = 0.00037$) were most variable between sites with presence and absence of *P. monophylla*. The variable slope was closely correlated with ruggedness as well as with MMAX and MTWM ($r_s > 0.7$). The p_{slope} of the Kruskal-Wallis test was larger than $p_{\text{ruggedness}}$ and p_{MMAX} was larger than p_{MTWM}. Slope and MMAX were therefore excluded from the multivariate model analysis. The stepwise multivariate binominal logistical and Random Forest models showed that the model for "presence of *P. monophylla*" included the independent variables ruggedness, ruggedness VRM and average temperature in the warmest month (MTWM) (Table 4).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

Table 5. Results of the multivariate binomial logistic regression model (AIC = 601.85; residual deviance = 593.85 on 588 degrees of freedom).

Factor	Estimate	RMSE	Z value	Pr(>|z|)
Intercept	26.38568	8.81813	2.992	0.00277
Ruggedness	0.18183	0.01579	11.519	<2e-16
MTWM	-1.19683	0.35920	-3.332	0.00086

The ruggedness factor was the most influential predictor variable and indicated that the probability of *P. monophylla* occurrence was larger than 50% when the degree of ruggedness was higher than 17.5 m (Fig. 4). The ruggedness VRM also indicated that a minimum change in roughness increases the probability of presence of the pine (Fig. 5). When MTWM increased from 23.5 to 25.2 °C, the probability of occurrence of *Pinus monophylla* decreased (Fig. 6). After cross validation (10-fold), the Random Forest model revealed that the variables ruggedness, ruggedness VRM and MTWM yielded a high correlation for their ability to predict presence of the *P. monophylla*; AUC = 0.920 TSS = 0.69, Kappa = 0.691, sensitivity was 0.812 and specificity was 0.878.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

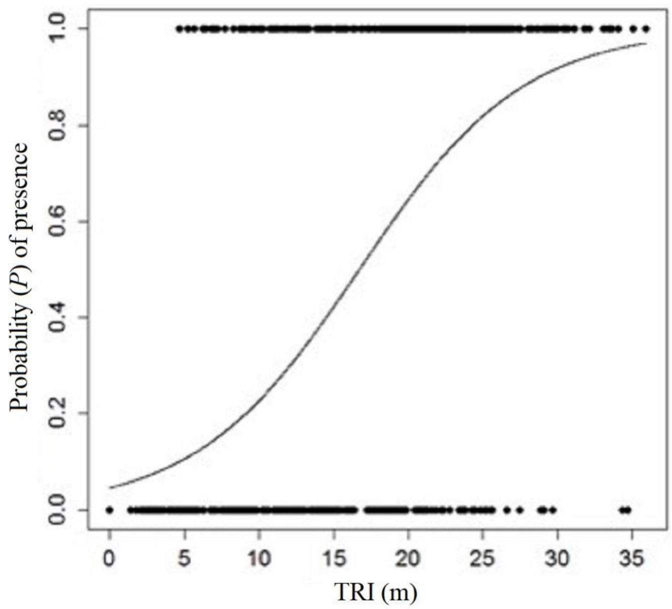


Figure 4. The relationship between the probability (*P*) of presence of *Pinus monophylla* and the terrain ruggedness index (TRI) (m) of the terrain in Sierra La Asamblea, Baja California, Mexico.

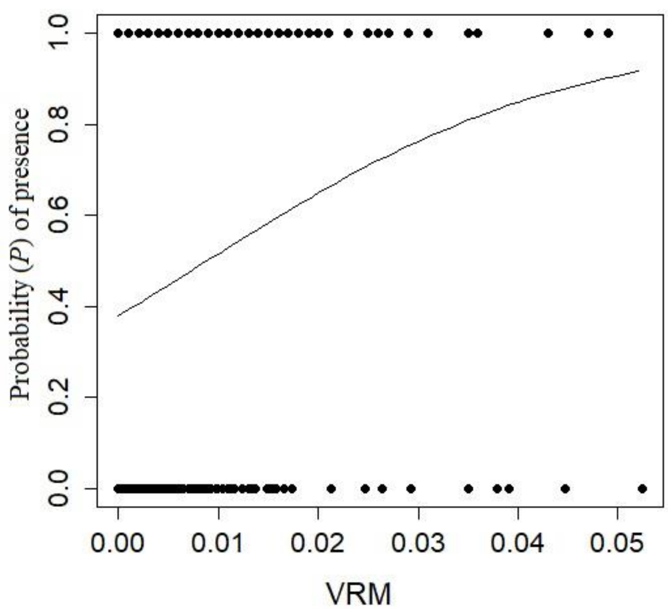


Figure 5. The relationship between the probability (*P*) of presence of *Pinus monophylla* and vector ruggedness measure (VRM) in Sierra La Asamblea, Baja California, Mexico.

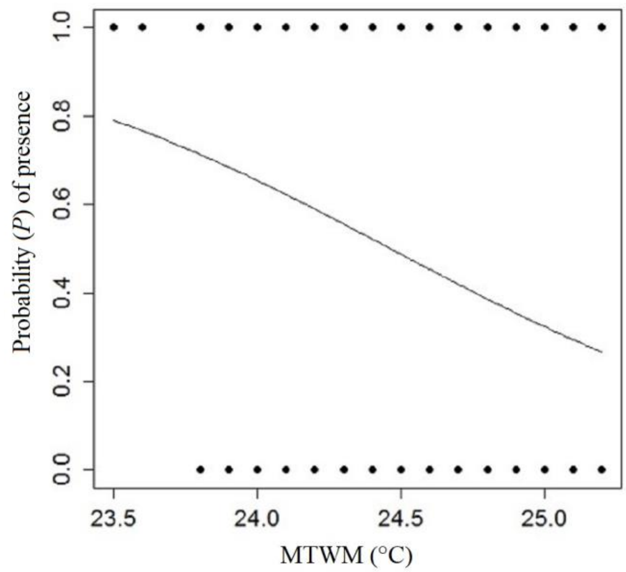


Figure 6. The relationship between the probability (*P*) of presence of *Pinus monophylla* and the average temperature in the warmest month (MTWM) in Sierra La Asamblea, Baja California, Mexico.

DISCUSSION

Pixel-based classification

Predicting the presence of pine forest by using BPNN proved feasible. The NDVI was one of the variables that contributed to the prediction and clearly separated forest cover (NDVI > 0.35) from the other types of vegetation cover (NDVI < 0.20). The overall accuracy of classification (*K* = 0.86) was similar to that reported in other studies using Sentinel-2A MSI images; for example, Immitzer et al., (2016) reported a *K* of 0.85 for tree prediction in Europe by using five classes and a random forest classifier. Vieira et al. (2003) reported a *K* = 0.77 in eastern Amazon using seven classes and 1999 Landsat 7 ETM imagery. However, Sothe et al. (2017) reported *K* values of 0.98 and 0.90 for respectively three successional forest stages and field in a subtropical forest in Southern Brazil by using Sentinel-2 and Landsat-8 data associated with the support vector machine algorithm. Kun et al. (2014) obtained *K* values of 0.70 to 0.85 for land-use type prediction (including forest) in China by using the support vector machine algorithm classifier and Landsat-8 images of lower spatial resolution than Sentinel images. The very high level of accuracy obtained by Kun et al. (2014) was probably due to the large-scale of the study and the clearly differentiated types of land considered.

Relationship between presence of *P. monophylla* and environmental variables

Ruggedness of the terrain was the most important topographic variable, significantly explaining the presence of pines in Sierra La Asamblea (Table 3). Ruggedness, which is strongly positively correlated with slope, may reduce solar radiation, air temperature and evapotranspiration due to increased shading (Tsujino et al., 2006; Bullock et al., 2008). The ruggedness indicated by the TRI index explains the presence of the pines because the Sierra La Asamblea is heterogeneous in terms of elevation. The VRM index was less important partly because the index is strongly dependent on the vector aspect (Gisbert & Martí, 2010) and in the case of the Asamblea the aspect is very homogeneous and the index values therefore tend to be very low (Fig. 5), as also reported by Wu et al. (2018). The pines were expected to colonize north facing slopes, which are exposed to less solar radiation than slopes facing other directions. However, the topographical variable aspect was not important in determining the presence of *P. monophylla* var. *californiarum* in the study site, possibly because of physiological adaptations regarding water-use efficiency and photosynthetic nitrogen-use efficiency (DeLucia & Schlesinger, 1991), as reported for the *Pinus monophylla, P. halepensis, P. edulis* and *P. remota* in arid zones (Lanner & Van Devender, 2000; Helman et al., 2017). The Mediterranean climate, with wet winters and dry summers, is another characteristic factor in this mountain range. In the winter in this part of the northern hemisphere, the sun, which is in a lower position and usually affects the southern aspect by radiation, was masked by clouds, rainfall and occasional snowfall (León-Portilla, 1988). During the summer, the level of solar radiation is greater, but similar in all directions because the sun is closest to its highest point (Stage & Salas, 2007).

The above-mentioned finding contrasts with those of other studies reporting that north-eastern facing slopes in the northern hemisphere receive less direct solar radiation, thus providing more favourable microclimatic conditions (air temperature, soil temperature, soil moisture) for forest development, permanence and productivity than southwest-facing sites (Astrom et al., 2007; Stage & Salas, 2007; Hang et al 2009; Marston et al., 2010; Klein et al., 2014). DeLucia & Schleinger (1991) reported that *P. monophylla* populations in the Great Basin California desert with summer rainfall (monsoon) preferred an east-southeast aspect with lower solar radiation and evapotranspiration.

The probability of presence of *P. monophylla* was also related to the climatic variable MTWM. In the Sierra La Asamblea, this pine species was found in a narrow range of MTWM of between 23.5° and 25.2° (Table 1), which, however, is a smaller range than reported for the other pine species (Tapias et al., 2004; Roberts & Ezcurra, 2012). Therefore, this species should adapt well to high temperatures in the summer (Lanner et al., 2000), which is usually a very dry period in the study site (León-Portilla, 1988). However, the probability of occurrence was greatest for an MTWM of 23.5°C (Fig. 5, which occurred at the top of the Sierra La Asamblea, at an elevation of about 1,660 m). We therefore conclude that this species can also grow well when the MTWM is below 23.5°C. On the other hand, considering MTWM as factor yielded a probability of occurrence of 25-80%. The spatial resolution of the climatic data by the national database run by

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

the University of Idaho is probably not adequate to describe the microhabitat of *P. monophylla* (Rehfeldt et al., 2006; Marston et al., 2010).

Identification of *P. monophylla* in the Sierra La Asamblea as the most southern populations represents an opportunity for research on climatic tolerance and community responses to climatic variation and change.

ACKNOWLEDGEMENTS

We are grateful to E. Espinoza, F. Macias and A. Guerrero for their support with the fieldwork.

REFERENCES

Abraira V. 2001. El índice kappa. *Semergen* 27:247-249. DOI:10.1016/S1138- 3593(01)73955-X

Allen CD, Macalady AK, Chenchouni H, Bachelet D, Vennetier M, Kitzberger G, Rigling H, Breshears D, Hoog T, Gonzalez PK., Fensham R, Zhangm Z, Castro J, Demidova N, Jong-Hwan L, Allard G, Running S, Semerci A, Cobbt N. 2010. A global overview of drought and heat-induced tree mortality reveals emerging climatic change risks for forest. *Forest ecology and management* 259:660-684. DOI: 10.1016/j.foreco.2009.09.001.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232. DOI:10.1111/j.1365-2664.2006.01214.x

Bailey DK. 1987. A study of *Pinus* subsection *Cembroides*. The single-needle pinyons of the Californias and the Great Basin. Notes from the Royal Botanic Garden, Edinburgh. 44:275-310.

Borràs J, Delegido J, Pezzola A, Pereira M, Morassi G, Camps-Valls G. 2017. Land use classification from Sentinel-2 imagery. *Revista de Teledetección* 48:55-66. DOI: 10.4995/raet.2017.7133.

Braspenning P J, Thuijsman F. 1995. Artificial neural networks: an introduction to ANN theory and practice. Springer Science & Business Media. USA. 295 p.

Brockmann Consult, 2017. Sentinel Application Platform (SNAP). Available at: http://step.esa.int/main. / (accessed 18 April 2017).

Bullock SH, Heath D. 2006. Growth rates and age of native palms in the Baja California desert. *Journal of Arid Environments* 67(3):391-402. DOI: 10.1016/j.jaridenv.2006.03.002.

Bullock SH, Salazar Ceseña JM, Rebman JP, Riemann H. 2008. Flora and vegetation of an isolated mountain range in the desert of Baja California. *The Southwestern Naturalist* 53:61-73. DOI: 10.1894/0038-4909(2008)53[61:FAVOAI]2.0.CO;2.

Callaway RM, DeLucia EH, Nowak R, Schlesinger WH. 1996. Competition and facilitation: contrasting effects of *Artemisia tridentata* on desert vs. montane pines. *Ecology* 77:2130-2141. DOI: 10.2307/2265707.

Chambers JC. 2001. *Pinus monophylla* establishment in an expanding *Pinus-Juniperus* woodland: Environmental conditions, facilitation and interacting factors. *Journal of Vegetation Science* 12:27-40.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

CONABIO. 2017. Comisión Nacional para el Conocimiento y uso de la Biodiversidad. Geoportal de información. Sistema Nacional de información sobre Biodiversidad. *Available at*: http://www.conabio.gob.mx/informacion/gis/ (accessed 12 February 2017).

Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment 37:35-46. DOI: 10.1016/0034-4257(91)90048-B

DeCastilho CV, Magnusson WE, de Araújo RNO, Luizao RC, Luizao FJ, Lima AP, Higuchi N. 2006. Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of soil and topography. *Forest ecology and management* 234:85-96. DOI: 10.1016/j.foreco.2006.06.024.

DeLucia, EH, & Schlesinger, WH. 1991. Resource-use efficiency and drought tolerance in adjacent Great Basin and sierran plants. *Ecology*, *72*(1), 51-58. DOI: 10.2307/1938901

Development Core Team. 2017. A language and environment for statistical computing. R foundation for statistical computing, Vienna Austria. *Available at*: http://www.R-project.org. (accessed 8 September 2017).

Drusch M, Del Bello U, Carlier S, Colin O., Fernández V, Gascón F, Hoersch B, Isola C, Laberinti, P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote sensing environment* 120:25-36. DOI: 10.1016/j.rse.2011.11.026.

Elliott KJ, Miniat CF, Pederson N, Laseter SH. 2005. Forest tree growth response to hydroclimate variability in the southern Appalachians. *Global Change Biology* 21(12):4627-4641. DOI: 10.1111/gcb.13045.

ESA, 2017. European Space Agency. Copernicus, Sentinel-2. *Available At*: http://www.esa.int (accessed 21 March 2016).

Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27:861–874. DOI: 10.1016/j.patrec.2005.10.010

García E. 1998. Clasificación de Köppen, modificado por García, E. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), 1998. Available at: http://www.conabio.gob.mx/informacion/gis/ (accessed 2 June 2017).

Gisbert FJG, Martí IC. 2010. Un índice de rugosidad del terreno a escala municipal a partir de Modelos de Elevación Digital de acceso público. *Documento de Trabajo*. Available at: https://wheui3.grupobbva.com/TLFU/dat/DT_7_2010.pdf

Goodchild MF. 1994. Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues. *Journal of Vegetation Science* 5:615-626. DOI: 10.2307/3235878.

Guo PT, Wu W, Sheng QK, Li MF, Liu HB, Wang ZY. 2013. Prediction of soil organic matter using artificial neural network and topographic indicators in hilly areas. *Nutrient cycling in agroecosystems* 95:333344. DOI: 10.1007/s10705-013-9566-9.

Helman D, Osem Y, Yakir D, Lensky IM. 2017. Relationships between climate, topography, water use and productivity in two key Mediterranean forest types with different water-use strategies. *Agricultural and Forest Meteorology* 232:319-330. DOI: 10.1016/j.agrformet.2016.08.018.

Hepner G, Logan T, Ritter N, Bryant N. 1990. Artificial neural network classification using a minimal training set. Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing* 56(4):469-473.

Immitzer M,Vuolo F, Atzberger C. 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing* 8:1-27. DOI: 10.3390/rs8030166.

INEGI. 2013. Conjunto de datos vectoriales de uso de suelo y vegetación escala 1:250 000, serie V. Instituto Nacional de Estadística y Geografía. Aguascalientes. *Available at*: http://www.conabio.gob.mx/informacion/gis/ (accessed 10 September 2015).

Klein T, Hoch G, Yakir D, Körner C. 2014. Drought stress, growth and nonstructural carbohydrate dynamics of pine trees in a semi-arid forest. *Tree physiology* 34:981-992. DOI: 10.1093/treephys/tpu071.

Kun J, Xiangqin W, Xingfa G, Yunjun J. Xianhong X, Bin L. 2014. Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto International* 29:941-951. DOI:10.1080/10106049.2014.894586.

Lanner RM, Van Devender TR. 2000. The recent history of pinyon pines. In: Richardson, D. M. (eds). The American Southwest, Cambridge University Press. 171–182

Léon-Portilla. 1988. Miguel del Barco, Historia natural y crónica de la antigua California. Universidad Nacional Autónoma de México, México. 483 p.

Madonsela S, Cho MA., Ramoelo A, Mutanga O. 2017. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing* 133: 116–127. DOI: 10.1016/j.isprsjprs.2017.10.008.

Marston, RA. 2010. Geomorphology and vegetation on hillslopes: interactions, dependencies, and feedback loops. *Geomorphology*, *116*(3-4), 206-217.

Moran RV. 1983. Relictual northern plants on peninsular mountain tops. In: Biogeography of the Sea of Cortez; University of California Press, Berkeley, USA. 408–410.

Olofsson O, Foody GM, Stehman SV, Woodcock CE. 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment* 129:122–131. DOI: 10.1016/j.rse.2012.10.031

Olofsson, P, Foody, GM, Herold, M, Stehman, SV, Woodcock, CE, Wulder, MA. 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42-57. DOI: 10.1016/j.rse.2014.02.015

Osem Y, Zangy E, Bney-Moshe E., Moshe Y, Karni N, Nisan Y. 2009. The potential of transforming simple structured pine plantations into mixed Mediterranean forests through natural regeneration along a rainfall gradient. *Forest Ecology Management* 259:14–23. DOI:10.1016/j.foreco.2009.09.034.

Pettorelli N. 2013. The Normalized Difference Vegetation Index. Oxford, University Press. United Kingdom. 194 p.

QGIS Development. 2016. QGIS Geographic Information System. Open source Geospatial Foundation. Available at: http//qgis.osgeo.org

Radoux J, Chomé G, Jacques DC, Waldner F, Bellemans N, Matton N, Lamarche C, d'Andrimont R, Defourny P. 2016. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sensing* 8(6):488. DOI:10.3390/rs8060488.

Rehfeldt GE. A spline model of climate for the Western United States. 2006. Gen Tech Rep. RMRS-GTR-165. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.

Rehfeldt GE, Crookston NL, Warwell MV, Evans JS. 2006. Empirical analyses of plant-climate relationships for the western United States. *International journal plant science* 167:1123–1150. DOI: 1058-5893/2006/16706-0005.

Richards JA. 1999. *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, p.240.

Riley SJ, Degloria SD, Elliot R. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*5:23–27 (http://arcscripts.esri.com/details.asp?dbid=12435).

Roberts N, Ezcurra E. Desert Climate. 2012. In: Rebman, JP, Roberts NC, ed. *Baja California Plant Field Guide*. San Diego Natural History Museum. San Diego, USA. 1-23.

Rouse JW, Haas RH, Schell A, Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, December 10–15 1974, Greenbelt, MD, NASA, Washington, DC, pp.301–317.

Sáenz-Romero C, Rehfeldt GE, Crookston NL, Duval P, St-Amant R, Beaulieu J, Richardson BA. 2010. Spline models of contemporary, 2030, 2060 and 2090 climates for Mexico and their use in understanding climate-change impacts on the vegetation. *Climatic Change*, 102:595-623. DOI:10.1007/s10584-009-9753-5.

Salas EAL, Valdez R, Michel S. 2017. Summer and winter habitat suitability of Marco Polo argali in southeastern Tajikistan: A modeling approach. *Heliyon* 3(11):e00445. DOI:10.1016/j.heliyon.2017.e00445.

Sappington, JM., Longshore, KM., Thompson, D. B. 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *Journal of wildlife management*, *71*(5):1419-1426. DOI: 10.2193/2005-723

Satage AR, Salas C. 2007. Interactions of Elevation, Aspect, and Slope in Models of Forest Species Composition and Productivity. *Forest Science* 53:486-492. Available at: http://www.ingentaconnect.com/

Silva-Flores R, Pérez-Verdín G, Wehenkel C. 2014. Patterns of tree species diversity in relation to climatic factors on the Sierra Madre Occidental, Mexico. *PloS one 9,* e105034. DOI: 10.1371/journal.pone.0105034.

Shirk AJ, Waring K, Cushman S, Wehenkel C, Leal-Sáenz A, Toney C, Lopez-Sanchez CA. 2017. Southwestern white pine (*Pinus strobiformis*) species distribution models predict large range shift and contraction due to climate change. *Forest Ecology Management* (in review).

SNAP. 2017. ESA´s Sentinel-toolbox ESA Sentinel Application Platform. Version 6.0.0.

Sothe C, Almeida CMD, Liesenberg V, Schimalski MB. 2017. Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. *Remote Sensing* 9(8):838. DOI:10.3390/rs9080838.

Spasojevic MJ, Bahlai CA, Bradley BA, Butterfield BJ, Tuanmu MN, Sistla S, Wiederholt R, Suding KN. 2016. Scaling up the diversity-resilience relationship with trait databases and remote sensing data: the recovery of productivity after wildfire. *Global Change Biology* 22(4):1421–1432. DOI: 10.1111/gcb.13174.

Tapias R, Climent J, Pardos JA, Gil L. 2004. Life histories of Mediterranean pines. *Plant Ecology* 171: 53-68. DOI:10.1023/B:VEGE.0000029383.72609.f0.

Telespazio VEGA Deutschland GmbH 2016. Sentinel-2 MSI-Level-2A. Prototype Processor Installation and User Manual. Available at: http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf

Tsujino R, Takafumi H, Agetsuma N, Yumoto T. 2006. Variation in tree growth, mortality and recruitment among topographic positions in a warm temperate forest. *Journal of*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018

Vegetation Science 17:281-290. DOI:10.1658/1100-9233(2006)17[281:VITGMA]2.0.CO;2.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S-Plus. Fourth Edition. New York, Springer.

Vieira ICG, de Almeida AS, Davidson EA, Stone TA, de Carvalho CJR, Guerrero JB. 2003. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. *Remote Sensing of Environment* 87(4):470-481. DOI:10.1016/j.rse.2002.09.002.

Wu W, Li AD, He XH, Ma R, Liu HB., Lv JK. 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture* 144:86-93. DOI: 10.1016/j.compag.2017.11.037.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v3 | CC BY 4.0 Open Access | rec: 15 Feb 2018, publ: 15 Feb 2018