Detection of *Pinus monophylla* forest in the Baja California desert by remote sensing

Jonathan G Escobar-Flores [1] , **Carlos A Lopez-Sanchez** [2] , **Sarahi Sandoval** [3] , **Marco A Marquez-Linares** [1] , **Christian Wehenkel** [Corresp. 4]

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

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

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

[4] Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Durango, Mexico

Corresponding Author: Christian Wehenkel
Email address: wehenkel@ujed.mx

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 in 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, 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 establish the distribution of *Pinus monophylla* var. *californiarum* in Sierra La Asamblea, Baja California (Mexico) using Sentinel-2 images, and ii) to test and describe the relationship between this distribution of *Pinus monophylla* and five topographic and 18 climate variables. We hypothesized that i) the Sentinel-2 images can be used to accurately detect the *P. monophylla* distribution in the study site due to higher 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.

Methods. It was used an atmospherically corrected a 12-bit Sentinel-2A MSI image with eleven spectral bands in the visible, near infrared, and short-wave infrared light region combined with the normalized differential vegetation index (NDVI). Supervised classification of this image was carried out using a backpropagation-type artificial neural network algorithm. Stepwise multivariate binominal logistical regression and Random Forest regression including cross valuation (10 fold) were used to model the associations between presence/absence of pines and the five topographical and 18 climate variables.

Results. Probably, *P. monophylla* covers 4,955 hectares in the isolated in Sierra La Asamblea, Baja California (Mexico) via supervised classification of Sentinel-2 satellite images. The NDVI was one of the variables that contributed to the detection and clearly separated the forest cover (NDVI > 0.35) from the other vegetation cover (NDVI < 0.20). The ruggedness was the best 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 (Kappa) was similar to 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 Sierra La Asamblea as the most southern populations represents an opportunity for research on climatic tolerance

and community responses to climatic variation and change.

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

Jonathan. G. Escobar-Flores [1], Carlos Antonio 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 in 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 establish the distribution of *P. monophylla* var. *californiarum* in Sierra La Asamblea, Baja California (Mexico) 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 detect the *P. monophylla* distribution in the study site due to higher 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. **Methods.** It was used an atmospherically corrected a 12-bit Sentinel-2A MSI image with eleven spectral bands in the visible, near infrared, and short-wave infrared light region combined with the normalized differential vegetation index (NDVI). Supervised classification of this image was carried out using a backpropagation-type artificial neural network algorithm. Stepwise multivariate binominal logistical regression and Random Forest regression 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.** Probably, *P. monophylla* covers 4,955 hectares in the isolated in Sierra La Asamblea via supervised classification of Sentinel-2 satellite images. The NDVI was one of the variables that contributed to the detection and clearly separated the forest cover (NDVI > 0.35) from the other vegetation cover (NDVI < 0.20). The ruggedness was the best 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 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 climatic variation and change.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017

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 in 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 and 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 importance of this relict pine species, its existence is not considered in most forest inventories in Mexico, and its distribution is generalized in vegetation cover maps (CONABIO, 2017).

Remote sensing techniques facilitate analysis of the temporal-space dynamics of the vegetation in isolated sites, as with the piñon pine in Sierra de la Asamblea. Tree species distribution is generally modulated by hydroclimatic variables and topographies (Elliot et al. 2005), and it is therefore possible to determine the spatial conditions that favour the presence of forests, especially by using digital terrain models (DTMs). Such models have shown, e.g., that tropical and temperate forests tend to grow faster and more densely in sites with variable elevation and slope (Decastilho et al. 2006; Spasojevic et al. 2016). Another attribute that can be analyzed using DTMs is the slope, aspect and terrain ruggedness index (Riley et al. 1999), used to express the difference in elevation of adjacent cells in a digital elevation grid. The less intense solar radiation to which northern orientations are exposed is known to promote the growth and productivity of vegetation (Osem et al. 2009). In addition, forests in less rugged sites (flatlands and valleys) are known to be particularly susceptible to fire, while heterogeneous and highly rugged forest sites are more dispersed and less likely to be affected by fire (Ganteaume and Jappiot, 2013). Geomorphic attributes such as ruggedness and profile convexity have recently been used to classify vegetation types from satellite images (Franklin et al. 2000; Carler & Wolff, 2004; Waser et al. 2011) and have been used along with multi-temporal analysis of images to identify forest types (Zhu & Liu, 2014).

Remote sensing with Landsat images has proved useful for detecting forest cover; the Landsat-8 satellite has sensors (7 bands) that can be used to analyze vegetation in spatial resolutions of 30 m (Johansen and Phinn, 2006). 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 meters per pixel) is three times higher than that of Landsat images, thus increasing their potential for detecting and differentiating between types of vegetation cover (Drush et al. 2012; Borras et al. 2017). The Sentinel-2 has 13 bands which 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 establish the distribution of *Pinus monophylla* var. *californiarum* in the Sierra La Asamblea, Baja California (Mexico) using Sentinel-2 images, and ii) to test and describe the relationship between this distribution of *Pinus monophylla* and five topographic and 18 climate variables. We hypothesized that i) the Sentinel-2 images can be used to accurately detect the *P. monophylla* distribution in the study site due to higher 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).

MATERIALS AND METHODS

Study area

Sierra La Asamblea is located in Baja California's central desert (-114° W 29° 19′ N, range of altitude 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. The 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 the 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).

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

Establishing the distribution of P. monophylla var. californiarum using Sentinel-2 images

Senitnel-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) Gloval 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 region with spatial resolutions of 10-60 m. However, the band one used for studies of coastal aerosol and the band ten applied for cirrus were not used in this study (ESA, 2017). Hence, the data preparation involved the resampling of the seven S2 bands acquired at 20 m and 60 m to obtain a layer stack of 11 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 to assess spectral indices with spatial reliability and products comparison, level-1C data have been converted to level-2A (top-of-canopy) taking into account the effects of aerosols and water vapor on reflectance (Radoux et al., 2016). These corrections have been realized using the Sea2Cor tool (Telespazio VEGA Deutschland GmbH, 2016) for the Sentinel-2 images.

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

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. 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 higher 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. Supervised classification of this image was carried out using a backpropagation-type artificial neural network (ANN) algorithm. The input weights corresponded to the values of the pixels twelve of Sentinel-2 and of the NDVI image. A logistic activation function was used with a training rate of 0.20 and 40 interactive processes. The network also calculates the error at each iteration (RMS) (Braspenning & Thuijisman, 1995). Additionally, the support vector machine algorithm (SVM) was used to classify the MSI (Mountrakis et al., 2011).

The training sites corresponded to two groups of georeferenced sites (Datum WGS-84, 11N) obtained during a project entitled “Oasis evaluation in Baja California”. The coordinates of the first group of sites whose were obtained in the field in October 2014 and October 2015. Four classes were defined with the object improving the discrimination between vegetation cover. The following classes were considered: i) pines, 502 sites, ii) scrub, 563 sites, iii) chaparral, 419 site and iv) no apparent vegetation, 419 sites. The second group comprised control sites and the coordinates were obtained by systematic sampling of the vegetation layers of the V series of land use and cover vegetation of the National Institute of Statistics and Geography (INEGI, 2015). The group included the following classes: i) pines, 596 sites, ii) scrub, 619 sites, iii) chaparral, 481 sites, and iv) no apparent vegetation, 418 sites.

The classification was validated using a confusion matrix, which is a table that compares the real values with the classification results. The confusion matrix was also used to determine the user accuracy, which refers to the total number of correct pixels/total number of reference pixels × 100% (Congalton, 1991). The accuracy of classification was calculated using the Kappa (*K*) coefficient. The *K* coefficient is a statistic used in accuracy assessment to measure whether one error matrix is significantly different from another. This statistic takes values of between -1 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).

Relationship between the distribution of P. monophylla and topographic and climate variables

To test and model the association between presence/absence of *P. monophylla* in the study area and topographical or climate variables, points estimates of the topographical variables ruggedness, slope, aspect, elevation and convexity and 18 climate variables (Table 2) were obtained from a national database managed by the University of Idaho (http://forest.moscowfsl.wsu.edu/climate/) and which requires point coordinates (latitude, longitude, and elevation) as the main inputs (Rehfeldt et al. 2006; Rehfeldt et al. 2006).

Table 2. Topographical and climatic variables considered in the study

For each variable in 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 ($\alpha = 0.0005$) were excluded from further analysis. The colinearity between the significant variables was measured using the Spearman correlation coefficient (r_s). When the r_s value for the difference between two significant variables was larger than 0.7, only the variable with the lowest p value in the Kruskal-Wallis test was used in the multivariate regressions. Finally, stepwise multivariate binominal logistical regression and Random Forest regression 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., 2017).

Regressions including cross valuation 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 regression models was evaluated using Akaike information criterion (*AIC*), root-mean-square error (*RMSE*) and pseudo coefficient of determination (R^2).

RESULTS

Our model showed a potential *P. monophylla* cover of 4,955 hectares in the in Sierra de la Asamblea, Baja California, Mexico. The supervised classification with ANN indicated an overall accuracy of identification 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 highest proportion of omission errors (27.65%) (Table 3, Fig. 2; Fig. 3). The value of NDVI in the pine forest fluctuated between 0.30 - 0.41, and in chaparral between 0.24 - 0.28. The lowest values of NDVI occurred in the scrub vegetation with values between 0.10 - 0.15. The analysis using the SVM classifier only showed overall accuracy of 72%.

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

Figure 2. (A) Detection of *Pinus monophylla* by neural network classification. The light yellow shading polygon represents pine forest published in the V series of INEGI (2013). (B) Distribution of pines in the rugged sites in the Sierra La Asamblea (Photograph by Jonathan Escobar).

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

The Kruskal-Wallis test indicated that the median values for ruggedness (p < 2.1e-16), slope (p < 2.2e-16), MTWM ($p = 0.000014$), MMAX ($p = 0.000048$) and SPRP ($p = 0.00037$) were most different 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 $p_{M\text{MAX}}$ larger than p_{MTWM}. Slope and MMAX were therefore excluded from the multivariate regression analysis. The stepwise multivariate binominal logistical and Random Forest regression showed that the best "presence of pines" model included the independent variables ruggedness and average temperature in the warmest month (MTWM) (Table 4).

Table 4. Results obtained with the best multivariate binomial logistic regression model (AIC = 611.96).

The ruggedness factor was the best 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 (Fig. 4). When MTWM increased from 23.5 to 25.2 °C, the probability of occurrence of *Pinus monophylla* decreased (Fig. 5). After cross validation (tenfold), the Random Forest model revealed that the variables ruggedness and MTWM explained the variation in the presence of *P. monophylla*, with $R^2 = 0.371$ and RMSE = 0.403.

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

Figure 5. The relationship between the probability (*P*) of occurrence of *Pinus monophylla* and the average temperature in the warmest month (MTWM) in Sierra La Asamblea, Baja California, Mexico.

DISCUSSION

Detection of pine forest by using ANN proved efficient. The NDVI was one of the variables that contributed to the detection and clearly separated the forest cover (NDVI > 0.35) from the other vegetation cover (NDVI < 0.20). The presence of the blue palm *Brahea armata* and fan palm *Washingtonia filifera*, which grow in the canyons and had values of NDVI greater than 0.30, may have confused the classification. However, these species are restricted to an elevation of less than 1,000 m, and were therefore excluded from the classification (Bullock et al. 2008).

The overall classification accuracy in this study ($K = 0.90$) was similar to other studies using Sentinel-2A MSI images. Immitzer et al. (2016) reported a K of 0.85 in tree detection in Europa using five classes and random forest classifier in Europa. Vieira et al. (2003) found a $K = 0.77$ in eastern Amazonia using seven classes and 1999 Landsat 7 ETM imagery. However, Sothe et al. (2017) reported a $K = 0.98$ and $K = 0.90$, respectively evaluating three successional forest stages and field in a subtropical forest in Southern Brazil by Sentinel-2 and Landsat-8 Data associated with the support vector machine algorithm. Kun et al. (2014) showed a K of 0.70 to 0.85 in land-use type detection including forests in China using the support vector machine algorithm classifier and Landsat-8 images providing lower spatial resolution than Sentinel. The cause of this very good accuracy of Kun et al. was probably the large-scale and clearly differentiated land-use types used as classes.

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 (Di Castri et al. 1981; Tsujino et al. 2006; Bullock et al. 2008).

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 and Schlesinger, 1991), as reported for the *Pinus monophylla, P. halepensis, P. edulis, 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 lower position of the sun, which normally affects stronger the southern aspect by radiation could not show to advantage due to clouds, rainfall and occasional snowfall (León-Portilla, 1988). During the summer, the level of solar radiation is high, but similar in all directions because the sun is closest to its highest point (Stage and 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 for the *P. monophylla* populations in the Great Basin California desert with summer rainfall (monsoon) that this tree species preferred an east-southeast aspect with lower solar radiation and evapotranspiration (DeLucia & Schleinger, 1991).

The probability of occurrence 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 wider temperature 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., 1998), which is usually a very dry period in the study site (León-Portilla, 1988). However, the probability of occurrence was the highest 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 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 in the field.

References

Abraira, V. 2001. El índice kappa. *Semergen 2*7: 247-249. DOI:10.1016/S1138- 3593(01)73955-X.

Allen, C.D., Macalady, A. .K., Chenchouni, H., Bachelet, D., Vennetier, M., Kitzberger, G., Rigling, H., Breshears, D., Hoog, T, Gonzalez, P. K., 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

Bailey, D. K. 1987. A study of Pinus subsection Cembroides. The single-needle pinyons of the California 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, S. H., Heath, D. 2006. Growth rates and age of native palms in the Baja California desert. *Journal of arid environments* 67:391–402- DOI: 10.1016/j.jaridenv.2006.03.002.

Bullock, S. H., Salazar Ceseña, J. M., Rebman, J. P., 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. R. M., DeLucia E. H., Nowak, R., Schlesinger W. H. 1996. Competition and facilitation: contrasting effects of *Artemisia tridentata* on desert vs. montane pines. *Ecology* 77*:*2130–2141. DOI: 10.2307/2265707

Carleer, A., Wolff, E. 2004. Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing* 70*:* 135–140. DOI: 10.14358/PERS.70.1.135

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

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, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment 37:35–46.

DeCastilho, C. V., Magnusson, W. E., de Araújo, R. N. O., Luizao, R. C., Luizao, F. J., Lima, A. P., 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

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, K.J., Miniat, C.F., Pederson, N., Laseter, S.H. 2005. Forest tree growth response to hydroclimate variability in the southern Appalachians. *Global Change Biology* 21: 4627–4641. DOI: 10.1111/gcb.13045

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

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* 21: 61–79. DOI: 10.1080/014311600210993

Ganteaume, A., Jappiot, M. 2013. What causes large fires in Southern France? *Forest Ecology and Management* 294: 76–85. DOI: 10.1016/j.foreco.2012.06.055

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).

Heithecker, T. D., Halpern, C. B. 2007. Edge-related gradients in microclimate in forest aggregates following structural retention harvests in western Washington. *Forest Ecology and Management* 48: 163–173. DOI: 10.1016/j.foreco.2007.05.003

Helman, D., Osem, Y., Yakir, D., and Lensky, I. M. 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

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. 2015. 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).

Johansen, K., Phinn, S. 2006. Mapping structural parameters and species composition of riparian vegetation using IKONOS and Landsat ETM+ data in Australian tropical savannahs. *Photogrammetric Engineering & Remote Sensing* 72: 71–80. DOI: 10.14358/PERS.72.1.71

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, R. M., Van Devender, T. R. 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.

Marston, R. A. 2010. Geomorphology and vegetation on hillslopes: Interactions, dependencies, and feedback loops. *Geomorphology* 116: 206–217. DOI: 10.1016/j.geomorph.2009.09.028

Mountrakis, G., Im, J. and Ogole, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3): 247-259. DOI: 10.1016/j.isprsjprs.2010.11.001

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

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.

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

Rehfeldt G. 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 G. E, Crookston N. L, Warwell M. V, Evans J. S. 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

Riley, S. J., S. D. Degloria, and R. Elliot. 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., y E. Ezcurra. Desert Climate. 2012. In: Rebman, J. P, Roberts. NC, eds. Baja California Plant Field Guide; San Diego Natural History Museum. San Diego, USA. 1–23

Satage, A.R., 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/

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

Sothe, C., Almeida, C.M.D., Liesenberg, V. and Schimalski, M.B. 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, M. J., Bahlai, C. A., Bradley, B. A., Butterfield, B.J., Tuanmu, M.N., Sistla, S., Wiederholt, R., Suding, K. N. 2016. Scaling up the diversity-resilience relationship with trait databases and remote sensing data: the recovery of productivity after wildfire. *Global Change Biology* 1421–1432. DOI: 10.1111/gcb.13174

Tapias R., Climent, J., Pardos., 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 Vegetation Science* 17:281-290 DOI:10.1658/1100-9233(2006)17[281:VITGMA]2.0.CO;2

Venables W. N, Ripley B. D. 2002. Modern Applied Statistics with S-Plus. Fourth Edition. New York, Springer.

Vieira, I.C.G., de Almeida, A.S., Davidson, E.A., Stone, T.A., de Carvalho, C.J.R. and Guerrero, J.B. 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

Waser, L. T., Ginzler, C., Kuechler, M., Baltsavias, E., Hurni, L. 2011. Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from Airborne Digital Sensor (ADS40) and RC30 data. *Remote Sensing of Environment* 115: 76–85. DOI: 10.1016/j.rse.2010.08.006

Zhu, X., Liu, D. 2014. Accurate mapping of forest types using dense seasonal Landsat time-series. *Journal of Photogrammetry and Remote Sensing* 96:1–11. DOI: 10.1016/j.isprsjprs.2014.06.012

Figure 1(on next page)

Map of Sierra La Asamblea.

The black circles indicate georeferenced sites occupied by *Pinus monophylla*

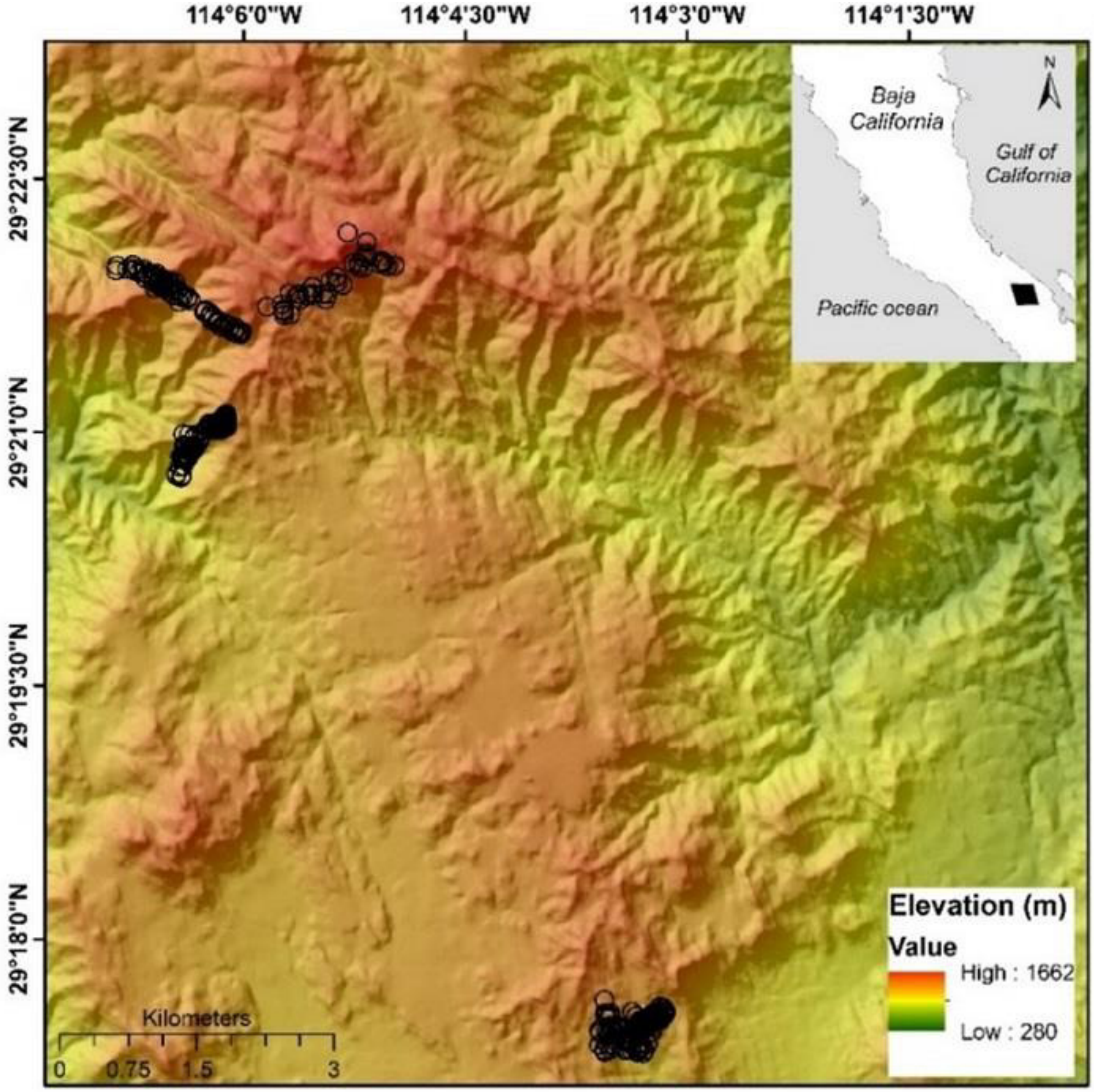
114°6'0"W
114°4'30"W
114°3'0"W
114°1'30"W
29°22'30"N
29°21'0"N
29°19'30"N
29°18'0"N
Baja California
Gulf of California
Pacific ocean
N
Elevation (m)
Value
High : 1662
Low : 280
Kilometers
0 0.75 1.5 3

Figure 2(on next page)

Detection of *Pinus monophylla*

(A) Detection of *Pinus monophylla* by neural network classification. The light yellow shading polygon represents pine forest published in the V series of INEGI (2013). (B) Distribution of pines in the rugged sites in the Sierra La Asamblea (Photograph by Jonathan Escobar)

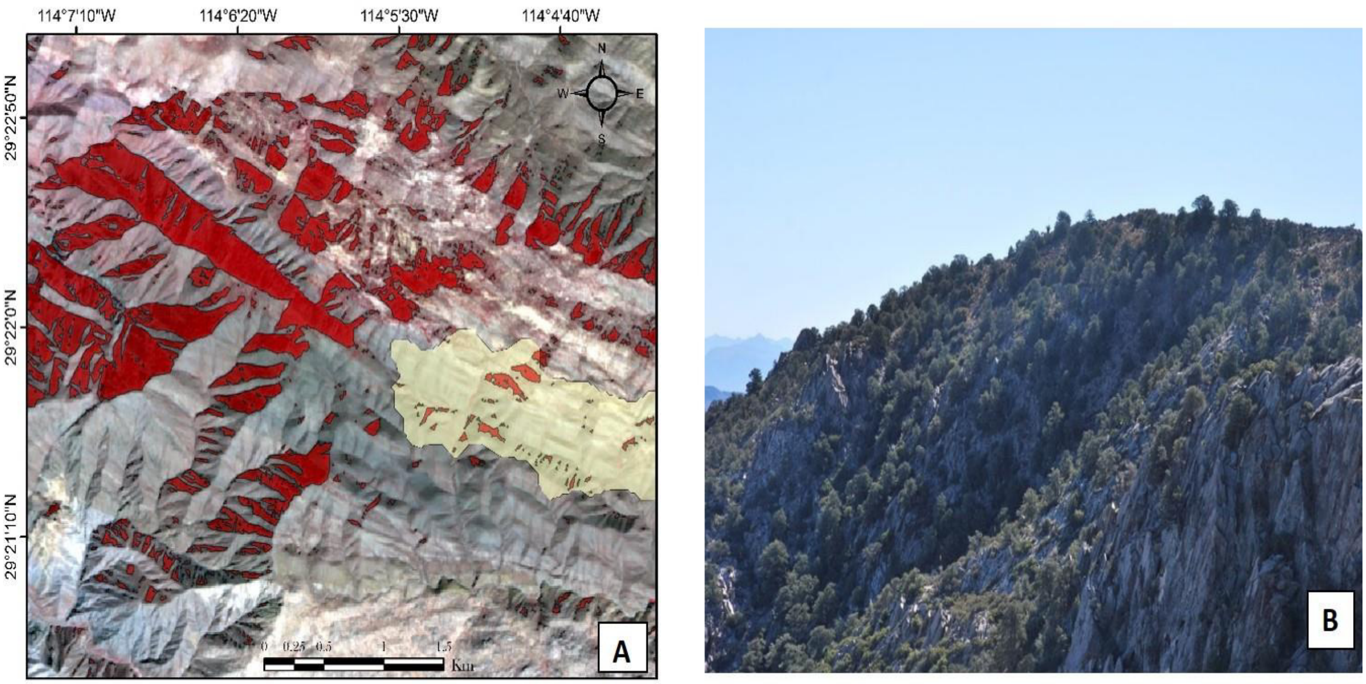
114°7'10"W
114°6'20"W
114°5'30"W
114°4'40"W
29°22'50"N
29°22'0"N
29°21'10"N
N
W
E
S
0 0.25 0.5 1 1.5
Km
A
B

Figure 3 (on next page)

Spectral signatures

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

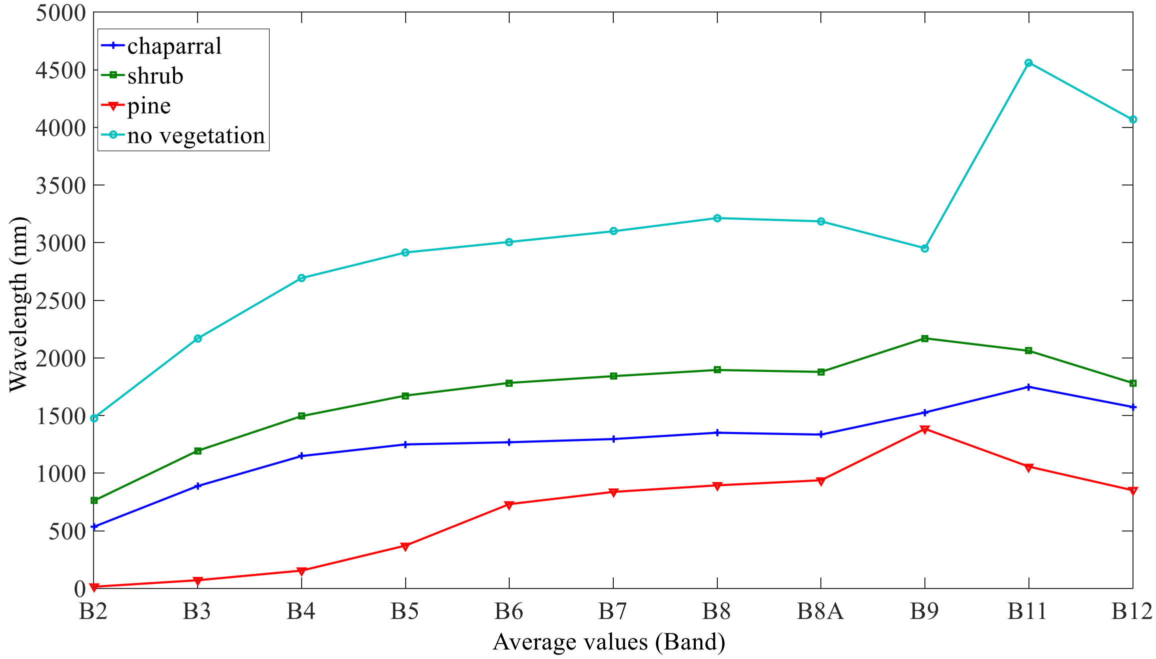
chaparral
shrub
pine
no vegetation
Wavelength (nm)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
B2
B3
B4
B5
B6
B7
B8
B8A
B9
B11
B12
Average values (Band)

Figure 4(on next page)

The relationship between the probability (*P*) of occurrence of *Pinus monophylla* and the ruggedness

The relationship between the probability (*P*) of occurrence of *Pinus monophylla* and the ruggedness (m) of the terrain in Sierra La Asamblea, Baja California, Mexico

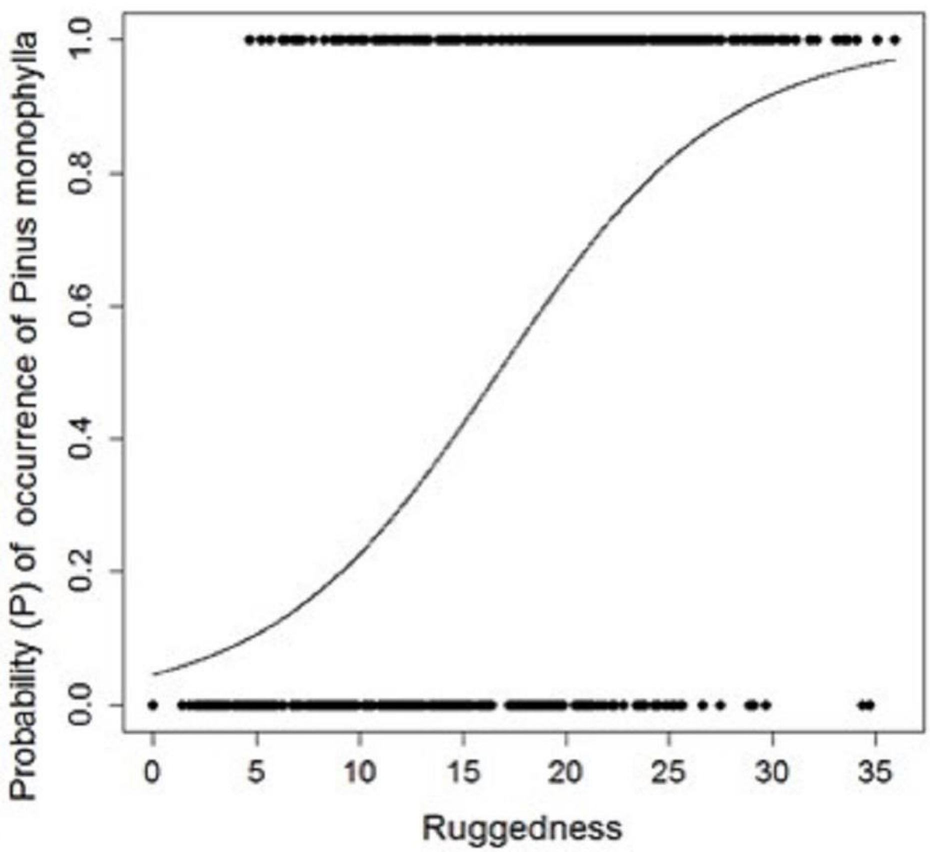
Probability (P) of occurrence of Pinus monophylla
Ruggedness
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35

Figure 5 (on next page)

The relationship between the probability (*P*) of occurrence of *Pinus monophylla* and the average temperature

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017

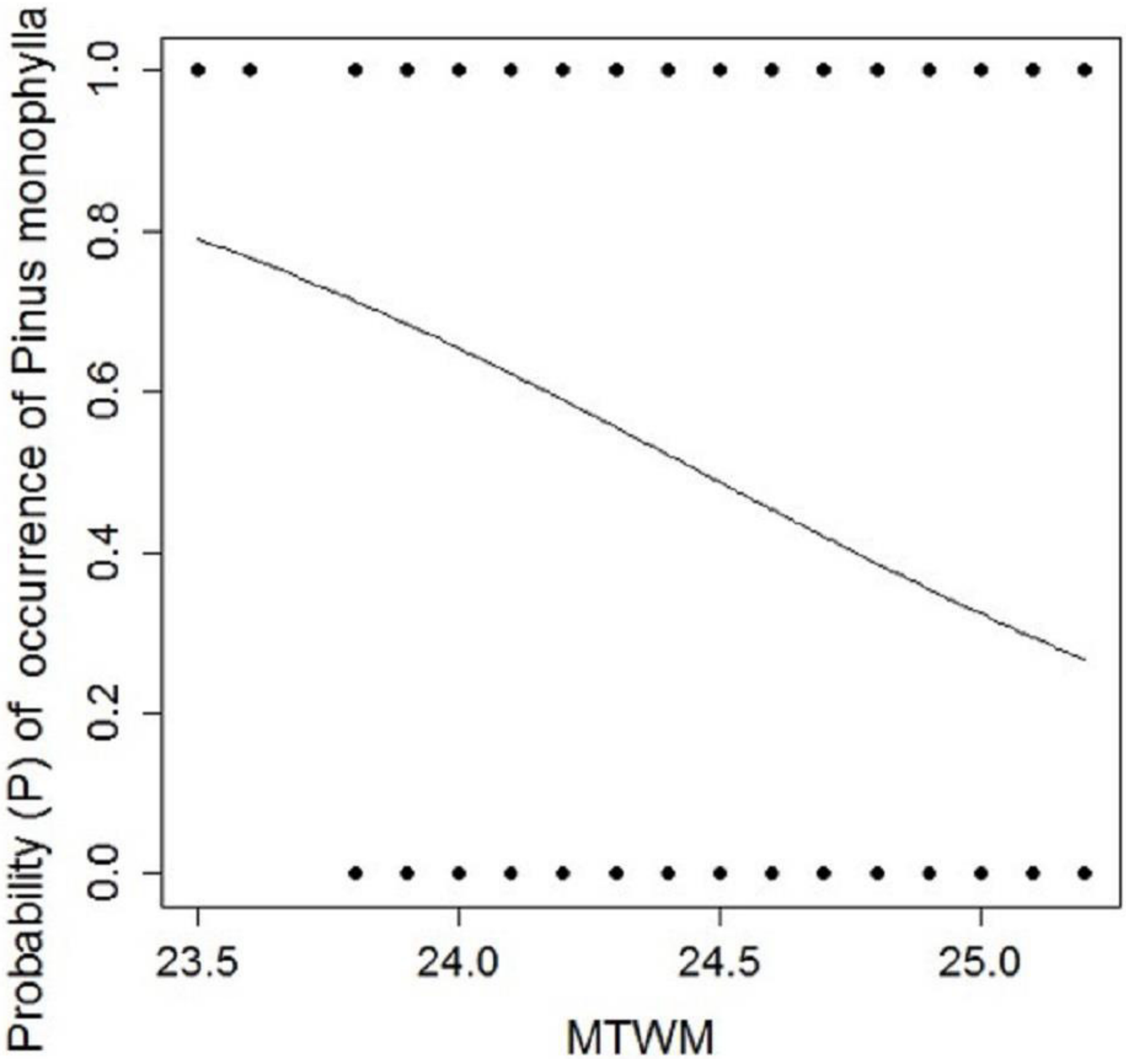
Probability (P) of occurrence of Pinus monophylla
1.0
0.8
0.6
0.4
0.2
0.0
23.5
24.0
24.5
25.0
MTWM

Table 1(on next page)

Sentinel-2 spectral bands

Sentinel-2 spectral bands used to detect the *Pinus monophylla* forest

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017

Bands	Central wave length (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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017

Table 2(on next page)

Topographical and climatic variables

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
Slope	S	º	28.38	8.92	48.34	3.42
Aspect	A	º	190.51	68.72	350.44	20.55
Elevation	E	m	1302.41	124.96	1631	1010
Convexity	C	º	-0.012	0.65	2.49	-2.44
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

Table 3 (on next page)

Results of the classification

Results of the classification monitored by neural network. The value of the Kappa coefficient was 0.862. The accuracy of classification of pine forest was 89.78%

Training set data (Known Cover Types) *						
Classification data	P	S	C	WV	Total	User accuracy (%)
P	522	0	14	0	536	87.58
S	24	619	119	2	764	100
C	50	0	348	7	405	72.35
WV	0	0	20	418	409	97.85
Total	596	619	481	418	2,114	

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

Table 4 (on next page)

Results obtained with the best multivariate binomial

Results obtained with the best multivariate binomial logistic regression model (AIC = 611.96)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3439v1 | CC BY 4.0 Open Access | rec: 29 Nov 2017, publ: 29 Nov 2017