A peer-reviewed version of this preprint was published in PeerJ on 4 April 2018.

<u>View the peer-reviewed version</u> (peerj.com/articles/4603), which is the preferred citable publication unless you specifically need to cite this preprint.

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 <u>https://doi.org/10.7717/peerj.4603</u>

Predicting *Pinus monophylla* forest cover 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

- 5 ¹ Instituto Politécnico Nacional. Centro Interdisciplinario De Investigación para el Desarrollo Integral
- 6 Regional, Unidad Durango., Durango, México
- 7 ² Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Durango,
- 8 México
- 9 ³ CONACYT Instituto Politécnico Nacional. CIIDIR. Unidad Durango. Durango, México
- 10
- 11 Corresponding author:
- 12 Christian Wehenkel²
- 13 Km 5.5 Carretera Mazatlán, Durango, 34120 Durango, México
- 14 Email address: <u>wehenkel@ujed.mx</u>
- 15
- 16

17 **ABSTRACT**

Background. The Californian single-leaf pinyon (Pinus monophylla var. californiarum), a 18 19 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 20 California (Mexico). This tree is distributed as a relict subspecies, at elevations of between 1,010 21 22 and 1,631 m in the geographically isolated arid Sierra La Asamblea (Baja California, Mexico), an area characterized by mean annual precipitation levels of between 184 and 288 mm. The aim of 23 this research was i) to estimate the distribution of P. monophylla var. californiarum in Sierra La 24 Asamblea by using Sentinel-2 images, and ii) to test and describe the relationship between the 25 distribution of *P. monophylla* and five topographic and 18 climate variables. We hypothesized that 26 i) Sentinel-2 images can be used to predict the *P. monophylla* distribution in the study site due to 27

28 the finer resolution (x_3) and greater number of bands (x_2) relative to Landsat-8 data, which is publically available free of charge and has been demonstrated to be useful for estimating forest 29 cover, and ii) the topographical variables aspect, ruggedness and slope are particularly important 30 because they represent important microhabitat factors that can determine the sites where conifers 31 can become established and persist. Methods. An atmospherically corrected a 12-bit Sentinel-2A 32 33 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 34 classification of this image was carried out using a backpropagation-type artificial neural network 35 36 algorithm (BPNN). Stepwise multivariate binominal logistical regression and Random Forest classification including cross valuation (10-fold) were used to model the associations between 37 presence/absence of *P. monophylla* and the five topographical and 18 climate variables. Results. 38 Using supervised classification of Sentinel-2 satellite images, we estimated that P. monophylla 39 covers $6,653 \pm 319$ hectares in the isolated Sierra La Asamblea. The NDVI was one of the variables 40 41 that contributed most to the prediction and clearly separated the forest cover (NDVI > 0.35) from the other vegetation cover (NDVI ≤ 0.20). Ruggedness was the most influential environmental 42 predictor variable, indicating that the probability of occurrence of *P. monophylla* was higher than 43 44 50% when the degree of ruggedness TRI was greater than 17.5 m. The probability of occurrence of the species decreased when the mean temperature in the warmest month increased from 23.5 to 45 25.2 °C. Discussion. The accuracy of classification was similar to that reported in other studies 46 47 using Sentinel-2A MSI images. Ruggedness is known to create microclimates and provides shade that minimizes evapotranspiration from pines in desert environments. Identification of the P. 48 49 monophylla stands in Sierra La Asamblea as the most southern populations represents an

opportunity for research on climatic tolerance and community responses to climate variability andchange.

52 INTRODUCTION

The Californian single-leaf pinyon (Pinus monophylla var. californiarum), a subspecies of the 53 single-leaf pinyon (the world's only 1-needled pine), inhabits semi-arid zones of the Mojave Desert 54 55 (southern Nevada and southeastern California, US) and also of northern Baja California (BC) (Mexico). It is both cold-tolerant and drought-resistant and is mainly differentiated from the typical 56 subspecies *Pinus monophylla* var. *monophylla* by a larger number of leaf resin canals and longer 57 fascicle-sheath scales (Bailey, 1987). This subspecies was first reported in BC in 1767 (Bullock et 58 al., 2006). The southernmost record of P. monophylla var. californiarum in America was 59 previously in BC, 26-30 miles north of Punta Prieta, at an elevation of 1,280 m (longitude -60 114°.155; latitude 29°.070, catalogue number ASU 0000235), and the type specimen is held in the 61 Arizona State University Vascular Plant Herbarium. 62

This tree is distributed as a relict subspecies 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) in areas with mean annual precipitation levels of between 184 and 288 mm (Roberts & Ezcurra, 2012). 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 of less than 150 mm. In fact, seeds of this variety survive well 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
(CONABIO, 2017).

Remote sensing with Landsat images has been demonstrated to be useful for estimating forest 76 77 cover; the The Landsat-8 satellite has sensors (7 bands) that can be used to analyze vegetation at a spatial resolution of 30 m (Madonsela et al., 2017). However, the European Space Agency's 78 Copernicus program has made Sentinel-2 satellite images available to the public free of charge. 79 80 The spatial resolution (10 m is pixel) of the images is three times finer that of Landsat images, thus increasing their potential for predicting and differentiating types of vegetation cover (Drush et al., 81 82 2012; Borras et al., 2017). The Sentinel-2 has 13 bands, of which 10 provide high-quality radiometric images of spatial resolution 10 to 20 m in the visible and infrared regions of the 83 electromagnetic spectrum. These images are therefore ideal for land classification (ESA, 2017). 84

The aim of this research was i) to estimate the distribution of *Pinus monophylla* var. californiarum 85 86 in 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 87 88 climate variables. We hypothesized that i) the Sentinel-2 images can be used to accurately predict 89 the *P. monophylla* distribution in the study site due to finer resolution (x3) and greater number of 90 bands (x2) than in Landsat-8 data, and ii) the topographical variables aspect, ruggedness and slope 91 are particularly influential because they represent important microhabitat factors that can 92 determine where conifers can become established and persist (Marston, 2010).

93 MATERIALS AND METHODS

94 Study area

Sierra La Asamblea is located in Baja California's central desert (-114° 9' W 29° 19' N, elevation 95 range 280-1,662 m, Fig. 1). The climate in the area is arid, with maximum temperatures of 40° C 96 in the summer (Garcia, 1998). The sierra is steeper on the western slopes, with an average incline 97 of 35°, and with numerous canyons with occasional springs and oases. Valleys and plateaus are 98 common in the proximity of the Gulf of California. Granite rocks occur south of the sierra and 99 meta-sedimentary rocks along the north and southeast of the slopes. The predominant type of 100 vegetation is xerophilous scrub, which is distributed at elevations ranging from 200 to 1,000 m. 101 102 Chaparral begins at an altitude of 800 m, and representative specimens of Adenostoma fasciculatum, Ambrosia ambrosioides, Dalea bicolor orcuttiana Quercus tuberculata, Juniperus 103 *california* and *Pinus monophylla* are also present at elevations above 1,000 m. Populations of the 104 endemic palm tree Brahea armata also occur in the lower parts of the canyons with superficial 105 water flow and through the rocky granite slopes (Bullock et al., 2006). 106

107

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

110



111

112 Datasets

113 Sentinel-2

114 The Sentinel-2A multispectral instrument (MSI) L1C dataset, acquired on 11 October 2016, in the trajectory of coordinates latitude 29°.814, longitude 114°.93, was downloaded from the US. 115 Geological Survey (USGS) Global Visualization Viewer at http://glovis.usgs.gov/. The 12-bit 116 117 Sentinel-2A MSI image has 13 spectral bands in the visible, NIR, and SWIR wavelength regions with spatial resolutions of 10-60 m. However, band one, used for studies of coastal aerosols, and 118 bands nine and ten, applied for respectively water vapour correction and cirrus detection, were not 119 120 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 10 121 m (Table 1) using the ESA's Sentinel-toolbox ESA Sentinel Application Platform (SNAP) and 122 then converted to ENVI format. 123

Because atmospherically improved images are essential to enable assessment of spectral indices with spatial reliability and product comparison, Level-1C data were converted to Level-2A

- 126 (Bottom of Atmosphere -BOA- reflectance) by taking into account the effects of aerosols and
- 127 water vapour on reflectance (Radoux et al., 2016). The corrections were made using the Sen2Cor
- tool (Telespazio VEGA Deutschland GmbH, 2016) for Sentinel-2 images.
- 129 Table 1. Sentinel-2 spectral bands used to predict the *Pinus monophylla* forest cover

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 11 –SWIR	1.610	20
Band 12 –SWIR	2.190	20

130

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 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.

138 Environmental variables

- 139 Tree species distribution is generally modulated by hydroclimate and topographical variables
- 140 (Elliot et al., 2005; Decastilho et al., 2006), which can be estimated from digital terrain models
- 141 (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 146 et al. (1999) and ii) a vector ruggedness measure (VRM), both implemented in QGIS (QGIS 147 Development Team, 2016). The TRI computes the values for each grid cell of a DEM. This 148 calculates the sum change in elevation between a grid cell and its eight-neighbor grid cell. VRM 149 incorporates the heterogeneity of both slope and aspect. This measure of ruggedness uses 3-150 151 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 152 al., 2007). 153

In addition, 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).

160

161

162

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	0	28.38	8.92	48.34	3.42
Aspect *	А	0	190.51	68.72	350.44	20.55
Elevation *	Е	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 $> 5^{\circ}$ C accumulating within the frost-free period *	GSDD5	Days	3491.82	164.76	3944	2995
Julian date when the sum degree days $> 5^{\circ}$ 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 2. Topographical and climatic variables considered in the study

164 * Variables for which no significant difference between the medians was obtained after

165 Bonferroni correction ($\alpha = 0.0005$) were excluded from further analysis.

166

167 **Pixel-based classification**

168 Classification method

Pixel-based classification was carried out in order to identify four different types of land cover in 169 the study area (P. monophylla, scrub, chaparral and no apparent vegetation). A supervised 170 171 classification approach with a backpropagation-type artificial neural network (BPNN) (SNAP, 2017) was applied. BPNN is widely used because of its structural simplicity and robustness in 172 modelling non-linear relationships. In this study, the BPNN comprises a set of three layers (raster): 173 174 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 175 the next layer (Guo et al., 2013). 176

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 predictions 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 QGIS software (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 estimated by an activation function. We used a logistic

function and training rate of 0.20, previously applied to land cover classification (Hepner et al., 190 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).

195 Validation

The BPNN classification was cross-validated (10-fold) using a confusion matrix, which is a table 196 that compares the reference data and the classification results. The confusion matrix was also used 197 to estimate the overall accuracy (the proportion of the area mapped correctly), user accuracy 198 199 (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 200 that category) (Congalton, 1991). We estimated the uncertainty of the classification through 201 202 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 estimate 203 204 the standard error, error-adjusted and 95% confidence intervals proposed by Olofsson et al. (2013). 205 The accuracy of classification was also estimated using the Kappa (K) coefficient. The K coefficient is often used as an overall measure of accuracy (Abraira, 2001). This coefficient takes 206 values of between 0 and 1, where values close to one indicate a high degree of agreement between 207 classes and observations, and a value of 0 suggests that the observed agreement is random (Abraira, 208 2001). However, the use of K is controversial because i) K would underestimate the probability 209 210 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). 211

212 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 213 and topographical or climate variables, a Kruskal-Wallis test was used to estimate the difference 214 in the median values in relation to presence and absence of *P. monophylla*. All variables for which 215 no significant difference between the median values was predicted after Bonferroni correction (α 216 = 0.0005) were excluded from further analysis. The collinearity between the variables with a 217 significant difference between the medians of presence and absence was estimated using the 218 219 Spearman correlation coefficient (r_s) . When the r_s value for the difference between two variables 220 was greater 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, 221 222 stepwise multivariate binominal logistical regression and Random Forest classification including 223 cross valuation (10-fold) were used to model the associations between presence/absence of P. 224 *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 the Akaike information criterion (AIC), root-mean-square error (RMSE) and residual deviance. Validation of the 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.

232 **RESULTS**

233 **Pixel-based classification**

234	We estimated the area of <i>P. monophylla</i> cover of $6,653 \pm 319$ hectares in Sierra de la Asamblea,
235	Baja California, Mexico. The supervised classification with BPNN yielded predictions with an
236	overall accuracy of identification of 87.74% (Table 3). This level of accuracy was estimated in the
237	32 interactions with 0.04 RMS training. The proportion of omission errors in the pine class was
238	only 12.42%, <i>i.e.</i> 87.58% of the pixels were correctly classified. The chaparral class had the larger
239	proportion of omission errors (27.65%) (Fig. 2; Fig. 3). The value of NDVI in the <i>P. monophylla</i>
240	forest fluctuated between 0.30 and 0.41, and in chaparral between 0.24 and 0.28. The lowest values
241	of NDVI corresponded to scrub vegetation, with values between 0.10 and 0.15.

Table 3. Estimated error matrix based of sample counts expressed as the estimated area

- 243 proportions (*W_i*). Accuracy measures are presented with a 95% confidence interval. Map
- 244 categories are the rows while the reference categories are the columns.

Classification data	Р	S	С	WV	Total	Wi	User´s	Producer's	Overall
Р	522	0	14	0	536	0.169	0.974±0.07	0.790±0.04	0.877±0.01
S	24	619	119	2	764	0.387	0.810±0.02	1.000	
С	50	0	348	7	405	0.258	0.859±0.01	0.752±0.07	
WV	0	0	20	418	438	0.186	0.954±0.002	0.970±0.02	
Total	596	619	501	427	2,143	1			

245 * P = *Pinus monophylla*; S = Shrub; C = Chaparral; WV= Without Vegetation; W_i = estimated

area proportions.

247

- **Figure 2**. (A) Estimated land cover classes using BPNN classification in Sierra La Asambla. (B)
- 249 Probability map of class assignment.



250

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



252

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

The Kruskal-Wallis test indicated that the median values for ruggedness TRI ($p < 2.1 \times 10^{-16}$), slope 254 $(p \le 2.2 \times 10^{-16})$, ruggedness VRM $(p = 4.9 \times 10^{-9})$, MTWM (p = 0.000014), MMAX (p = 0.000048)255 and SPRP (p = 0.00037) were most variable between sites in which *P. monophylla* was present 256 and absent. The variable slope was closely correlated with ruggedness as well as with MMAX and 257 MTWM ($r_s > 0.7$). The p_{slope} of the Kruskal-Wallis test was larger than $p_{\text{ruggedness}}$ and p_{MMAX} was 258 larger than p_{MTWM} . Slope and MMAX were therefore excluded from the multivariate model 259 260 analysis. The stepwise multivariate binominal logistical and Random Forest models showed that the "presence of P. monophylla" model included the independent variables ruggedness, ruggedness 261 VRM and average temperature in the warmest month (MTWM) (Table 4). 262

The ruggedness factor was the most influential predictor variable and indicated that the probability 263 264 of *P. monophylla* occurrence was larger than 50% when the degree of ruggedness TRI was higher than 17.5 m (Table 4). The ruggedness VRM also indicated that a minimum change in roughness 265 increases the probability of presence of the pine. The probability of occurrence of Pinus 266 monophylla decreased when MTWM increased from 23.5 to 25.2 °C (Table 5). After cross 267 validation (10-fold), the Random Forest model revealed that the variables ruggedness TRI, 268 ruggedness VRM and MTWM yielded a high correlation for their ability to predict presence of the 269 *P. monophylla* (AUC = 0.920, TSS = 0.69, Kappa = 0.691). The sensitivity was 0.812 and 270 specificity was 0.878. 271

Table 4. Results of the multivariate binomial logistic regression model (AIC = 601.8; residual
deviance= 593.85 on 588 degrees of freedom), TRI = terrain ruggedness index, VRM = vector
ruggedness measure, MTWM = mean temperature in the warmest month.

Variable	Estimate	Std. Error	Z value	Pr(> z)
Intercept	25.351	8.895	2.850	0.0044
MTWM	-1.159	0.362	-3.201	0.0014
Roughness TRI	0.178	0.015	11.200	< 2e-16
Roughness VRM	28.476	13.847	2.056	0.0397

275

276 **DISCUSSION**

277 **Pixel-based classification**

278 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 279 the other types of vegetation cover (NDVI < 0.20). The overall accuracy of classification (K =280 0.87) was similar to that reported in other studies using Sentinel-2A MSI images; for example, 281 282 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 283 classes and 1999 Landsat 7 ETM imagery. However, Sothe et al. (2017) reported K values of 0.98 284 285 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 286 algorithm. Kun et al. (2014) estimated K values of 0.70 to 0.85 for land-use type prediction 287 (including forest) in China by using the support vector machine algorithm classifier and Landsat-288 8 images of rougher spatial resolution than Sentinel images. The very high accuracy of predictions 289 290 by Kun et al. (2014) was probably due to the large-scale of the study and the clearly differentiated types of land considered. 291

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

Ruggedness of the terrain was the most important topographic variable, significantly explaining 293 the presence of pines in Sierra La Asamblea (Table 3). Ruggedness, which is strongly positively 294 correlated with slope, may reduce solar radiation, air temperature and evapotranspiration due to 295 increased shading (Tsujino et al., 2006; Bullock et al., 2008). The ruggedness indicated by the TRI 296 index explains the presence of the pines because Sierra La Asamblea is heterogeneous in terms of 297 elevation. The VRM index was less important partly because the index is strongly dependent on 298 299 the vector aspect (Gisbert & Martí, 2010) and in the case of Sierra Asamblea the aspect is very homogeneous and the index values therefore tend to be very low (Table 4), as also reported by Wu 300 et al. (2018). The pines were expected to colonize north facing slopes, which are exposed to less 301 302 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, 303 possibly because of physiological adaptations regarding water-use efficiency and photosynthetic 304 nitrogen-use efficiency (DeLucia & Schlesinger, 1991), as reported for the Pinus monophylla, P. 305 halepensis, P. edulis and P. remota in arid zones (Lanner & Van Devender, 2000; Helman et al., 306 2017). The Mediterranean climate, with wet winters and dry summers, is another characteristic 307 factor in this mountain range. In the winter in this part of the northern hemisphere, the sun (which 308 is in a lower position and usually affects the southern aspect by radiation) is masked by clouds, 309 310 rainfall and occasional snowfall (León-Portilla, 1988). During the summer, the solar radiation is more intense, but similar in all directions because the sun is closest to its highest point (Stage & 311 Salas, 2007). 312

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 less intense solar radiation and
evapotranspiration.

321 The probability of occurrence of *P. monophylla* was also related to the climatic variable MTWM. In Sierra La Asamblea, this pine species was found in a narrow range of MTWM of between 23.5° 322 and 25.2° (Table 1), which, however, is a smaller range than reported for the other pine species 323 324 (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 325 326 site (León-Portilla, 1988). However, the probability of occurrence was greatest for an MTWM of 327 23.5°C (Table 4), which occurred at the top of 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. 328 On the other hand, considering MTWM as factor yielded a probability of occurrence of 25-80%. 329 The spatial resolution of the climatic data by the national database run by the University of Idaho 330 is probably not adequate for describing the microhabitat of *P. monophylla* (Rehfeldt et al., 2006; 331 332 Marston et al., 2010).

Identification of the *P. monophylla* stands 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.

336 ACKNOWLEDGEMENTS

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

338 **REFERENCES**

- 339 Abraira V. 2001. El índice kappa. Semergen 27:247-249. DOI:10.1016/S1138-3593(01)73955-X
- 340 Allen CD, Macalady AK, Chenchouni H, Bachelet D, Vennetier M, Kitzberger G, Rigling H,
- 341 Breshears D, Hoog T, Gonzalez PK., Fensham R, Zhangm Z, Castro J, Demidova N, Jong-
- 342 Hwan L, Allard G, Running S, Semerci A, Cobbt N. 2010. A global overview of drought
- 343 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. 44:275350 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.

356	Brockmann Consult, 2017. Sentinel Application Platform (SNAP). Available at:
357	http://step.esa.int/main. / (accessed 18 April 2017).
358	Bullock SH, Heath D. 2006. Growth rates and age of native palms in the Baja California desert.
359	Journal of Arid Environments 67(3):391-402. DOI: 10.1016/j.jaridenv.2006.03.002.
360	Bullock SH, Salazar Ceseña JM, Rebman JP, Riemann H. 2008. Flora and vegetation of an isolated
361	mountain range in the desert of Baja California. The Southwestern Naturalist 53:61-73. DOI:
362	10.1894/0038-4909(2008)53[61:FAVOAI]2.0.CO;2.
363	Callaway RM, DeLucia EH, Nowak R, Schlesinger WH. 1996. Competition and facilitation:
364	contrasting effects of Artemisia tridentata on desert vs. montane pines. Ecology 77:2130-
365	2141. DOI: 10.2307/2265707.
366	Chambers JC. 2001. Pinus monophylla establishment in an expanding Pinus-Juniperus woodland:
367	Environmental conditions, facilitation and interacting factors. Journal of Vegetation Science
368	12:27-40.
369	CONABIO. 2017. Comisión Nacional para el Conocimiento y uso de la Biodiversidad. Geoportal
370	de información. Sistema Nacional de información sobre Biodiversidad. Available at:
371	http://www.conabio.gob.mx/informacion/gis/ (accessed 12 February 2017).
372	Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed
373	data. Remote sensing of environment 37:35-46. DOI: 10.1016/0034-4257(91)90048-B
374	DeCastilho CV, Magnusson WE, de Araújo RNO, Luizao RC, Luizao FJ, Lima AP, Higuchi N.
375	2006. Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of

376	soil and topography. Forest ecology and management 234:85-96. DOI:
377	10.1016/j.foreco.2006.06.024.
378	DeLucia, EH, & Schlesinger, WH. 1991. Resource-use efficiency and drought tolerance in
379	adjacent Great Basin and sierran plants. <i>Ecology</i> , 72(1), 51-58. DOI: 10.2307/1938901
380	Development Core Team. 2017. A language and environment for statistical computing. R
381	foundation for statistical computing, Vienna Austria. Available at: http://www.R-
382	project.org. (accessed 8 September 2017).
383	Drusch M, Del Bello U, Carlier S, Colin O., Fernández V, Gascón F, Hoersch B, Isola C, Laberinti,
384	P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P. 2012. Sentinel-2:
385	ESA's Optical High-Resolution Mission for GMES Operational Services. Remote sensing
386	environment 120:25-36. DOI: 10.1016/j.rse.2011.11.026.
387	Elliott KJ, Miniat CF, Pederson N, Laseter SH. 2005. Forest tree growth response to hydroclimate
388	variability in the southern Appalachians. Global Change Biology 21(12):4627-4641. DOI:
389	10.1111/gcb.13045.
390	ESA, 2017. European Space Agency. Copernicus, Sentinel-2. Available At: http://www.esa.int
391	(accessed 21 March 2016).
392	Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27:861–874. DOI:
393	10.1016/j.patrec.2005.10.010

394	García E. 1998. Clasificación de Köppen, modificado por García, E. Comisión Nacional para el
395	Conocimiento y Uso de la Biodiversidad (CONABIO), 1998. Available at:
396	http://www.conabio.gob.mx/informacion/gis/ (accessed 2 June 2017).
397	Gisbert FJG, Martí IC. 2010. Un índice de rugosidad del terreno a escala municipal a partir de
398	Modelos de Elevación Digital de acceso público. Documento de Trabajo. Available at:
399	https://wheui3.grupobbva.com/TLFU/dat/DT_7_2010.pdf
400	Goodchild MF. 1994. Integrating GIS and remote sensing for vegetation analysis and modeling:
401	methodological issues. Journal of Vegetation Science 5:615-626. DOI: 10.2307/3235878.
402	Guo PT, Wu W, Sheng QK, Li MF, Liu HB, Wang ZY. 2013. Prediction of soil organic matter
403	using artificial neural network and topographic indicators in hilly areas. Nutrient cycling in
404	agroecosystems 95:333344. DOI: 10.1007/s10705-013-9566-9.
405	Helman D, Osem Y, Yakir D, Lensky IM. 2017. Relationships between climate, topography, water
406	use and productivity in two key Mediterranean forest types with different water-use
407	strategies. Agricultural and Forest Meteorology 232:319-330. DOI:
408	10.1016/j.agrformet.2016.08.018.
409	Hepner G, Logan T, Ritter N, Bryant N. 1990. Artificial neural network classification using a
410	minimal training set. Comparison to conventional supervised classification.
411	Photogrammetric Engineering and Remote Sensing 56(4):469-473.
412	Immitzer M, Vuolo F, Atzberger C. 2016. First Experience with Sentinel-2 Data for Crop and Tree
413	Species Classifications in Central Europe. <i>Remote Sensing</i> 8:1-27. DOI: 10.3390/rs8030166.

NOT PEER-REVIEWED

414	INEGI. 2013. Conjunto de datos vectoriales de uso de suelo y vegetación escala 1:250 000, serie
415	V. Instituto Nacional de Estadística y Geografía. Aguascalientes. Available at:
416	http://www.conabio.gob.mx/informacion/gis/ (accessed 10 September 2015).
417	Klein T, Hoch G, Yakir D, Körner C. 2014. Drought stress, growth and nonstructural carbohydrate
418	dynamics of pine trees in a semi-arid forest. Tree physiology 34:981-992. DOI:
419	10.1093/treephys/tpu071.
420	Kun J, Xiangqin W, Xingfa G, Yunjun J. Xianhong X, Bin L. 2014. Land cover classification
421	using Landsat 8 Operational Land Imager data in Beijing, China. Geocarto International
422	29:941-951. DOI:10.1080/10106049.2014.894586.
423	Lanner RM, Van Devender TR. 2000. The recent history of pinyon pines. In: Richardson, D. M.
424	(eds). The American Southwest, Cambridge University Press. 171–182
425	Léon-Portilla. 1988. Miguel del Barco, Historia natural y crónica de la antigua California.
426	Universidad Nacional Autónoma de México, México. 483 p.
427	Madonsela S, Cho MA., Ramoelo A, Mutanga O. 2017. Remote sensing of species diversity using
428	Landsat 8 spectral variables. ISPRS Journal of Photogrammetry and Remote Sensing 133:
429	116–127. DOI: 10.1016/j.isprsjprs.2017.10.008.
430	Marston, RA. 2010. Geomorphology and vegetation on hillslopes: interactions, dependencies, and
431	feedback loops. Geomorphology, 116(3-4), 206-217.
432	Moran RV. 1983. Relictual northern plants on peninsular mountain tops. In: Biogeography of the
433	Sea of Cortez; University of California Press, Berkeley, USA. 408–410.

434	Olofsson O, Foo	ody GM, Stehn	man SV, W	oodcock C	CE. 20	013. Making bet	ter use of accurat	cy data
435	in land ch	ange studies:	Estimating	g accuracy	and	area and quanti	fying uncertainty	[,] using
436	stratified	estimation.	Remote	Sensing	of	Environment	129:122–131.	DOI:
437	10.1016/j.:	rse.2012.10.03	31					

- 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.
- 447 QGIS Development. 2016. QGIS Geographic Information System. Open source Geospatial
 448 Foundation. Available at: http://qgis.osgeo.org
- 449 Radoux J, Chomé G, Jacques DC, Waldner F, Bellemans N, Matton N, Lamarche C, d'Andrimont
- 450 R, Defourny P. 2016. Sentinel-2's potential for sub-pixel landscape feature detection.
- 451 *Remote Sensing* 8(6):488. DOI:10.3390/rs8060488.

452	Rehfeldt GE. A spline model of climate for the Western United States. 2006. Gen Tech Rep.
453	RMRS-GTR-165. U.S. Department of Agriculture, Forest Service, Rocky Mountain
454	Research Station, Fort Collins, Colorado, USA.
455	Rehfeldt GE, Crookston NL, Warwell MV, Evans JS. 2006. Empirical analyses of plant-climate
456	relationships for the western United States. International journal plant science 167:1123-
457	1150. DOI: 1058-5893/2006/16706-0005.
458	Richards JA. 1999. Remote Sensing Digital Image Analysis, Springer-Verlag, Berlin, p.240.
459	Riley SJ, Degloria SD, Elliot R. 1999. A terrain ruggedness index that quantifies topographic
460	heterogeneity. Intermountain Journal of Sciences5:23-27
461	(http://arcscripts.esri.com/details.asp?dbid=12435).
462	Roberts N, Ezcurra E. Desert Climate. 2012. In: Rebman, JP, Roberts NC, ed. Baja California
463	Plant Field Guide. San Diego Natural History Museum. San Diego, USA. 1-23.
464	Rouse JW, Haas RH, Schell A, Deering DW. 1974. Monitoring vegetation systems in the Great
465	Plains with ERTS. Proceedings of the Third Earth Resources Technology Satellite-1
466	Symposium, December 10–15 1974, Greenbelt, MD, NASA, Washington, DC, pp.301–317.
467	Sáenz-Romero C, Rehfeldt GE, Crookston NL, Duval P, St-Amant R, Beaulieu J, Richardson BA.
468	2010. Spline models of contemporary, 2030, 2060 and 2090 climates for Mexico and their
469	use in understanding climate-change impacts on the vegetation. Climatic Change, 102:595-
470	623. DOI:10.1007/s10584-009-9753-5.

471	Salas EAL, Valdez R, Michel S. 2017. Summer and winter habitat suitability of Marco Polo argali
472	in southeastern Tajikistan: A modeling approach. Heliyon 3(11):e00445.
473	DOI:10.1016/j.heliyon.2017.e00445.
474	Sappington, JM., Longshore, KM., Thompson, D. B. 2007. Quantifying landscape ruggedness for
475	animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. Journal of
476	wildlife management, 71(5):1419-1426. DOI: 10.2193/2005-723
477	Satage AR, Salas C. 2007. Interactions of Elevation, Aspect, and Slope in Models of Forest Species
478	Composition and Productivity. <i>Forest Science</i> 53:486-492. Available at:
479	http://www.ingentaconnect.com/
480	Silva-Flores R, Pérez-Verdín G, Wehenkel C. 2014. Patterns of tree species diversity in relation
481	to climatic factors on the Sierra Madre Occidental, Mexico. PloS one 9, e105034. DOI:
482	10.1371/journal.pone.0105034.
483	Shirk AJ, Waring K, Cushman S, Wehenkel C, Leal-Sáenz A, Toney C, Lopez-Sanchez CA. 2017.
484	Southwestern white pine (Pinus strobiformis) species distribution models predict large range
485	shift and contraction due to climate change. Forest Ecology Management (in review).
486	SNAP. 2017. ESA's Sentinel-toolbox ESA Sentinel Application Platform. Version 6.0.0.
487	Sothe C, Almeida CMD, Liesenberg V, Schimalski MB. 2017. Evaluating Sentinel-2 and Landsat-
488	8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil. Remote
489	Sensing 9(8):838. DOI:10.3390/rs9080838.

490	Spasojevic MJ, Bahlai CA, Bradley BA, Butterfield BJ, Tuanmu MN, Sistla S, Wiederholt R,
491	Suding KN. 2016. Scaling up the diversity-resilience relationship with trait databases and
492	remote sensing data: the recovery of productivity after wildfire. Global Change Biology
493	22(4):1421–1432. DOI: 10.1111/gcb.13174.
494	Tapias R. Climent J. Pardos JA, Gil L. 2004. Life histories of Mediterranean pines. <i>Plant Ecology</i>
405	171: 52 68 DOI:10 1022/P:VEGE 0000020383 72600 f0
495	171. 55-08. DOI.10.1025/D. VEGE.0000027585.72007.10.
496	Telespazio VEGA Deutschland GmbH 2016. Sentinel-2 MSI-Level-2A. Prototype Processor
497	Installation and User Manual. Available at:
498	http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf
499	Tsujino R, Takafumi H, Agetsuma N, Yumoto T. 2006. Variation in tree growth, mortality and
500	recruitment among topographic positions in a warm temperate forest. Journal of
501	Vegetation Science 17:281-290. DOI:10.1658/1100-
502	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.

- 509 Wu W, Li AD, He XH, Ma R, Liu HB., Lv JK. 2018. A comparison of support vector machines,
- 510 artificial neural network and classification tree for identifying soil texture classes in
- 511 southwest China. *Computers and Electronics in Agriculture* 144:86-93. DOI:
- 512 10.1016/j.compag.2017.11.037.