NORDIC JOURNAL OF
BOTANY

Research

Influence of physiography, soil and climate on *Taxus globosa*

Pablo Antúnez

P. Antúnez (https://orcid.org/0000-0002-5492-0836) ✉ (pantunez4@gmail.com), Forest Inventory and Remote Sensing, Faculty of Forest Sciences, Univ. of Goettingen, Göttingen, Germany, and: División de Estudios de Postgrado, Univ. de la Sierra Juárez, Oaxaca, México.

Nordic Journal of Botany
2021: e03058
doi: 10.1111/njb.03058

Subject Editor:
Jean Wan Hong Yong
Editor-in-Chief: Sara Cousins
Accepted 2 February 2021
Published 12 March 2021

In order to apply, promote and consolidate the actions and strategies of conservation of a species in danger of extinction, it is desirable to have the most detailed information possible about its habitat and the variables that affect it. In this study, we examined how temperature, soil pH, precipitation and physiography affect occurrence of *Taxus globosa*, an endangered species in Central and North America. The medical importance (contains taxol, a substance to treat cancer) of this plant and its endemic status make it a highly priority species that requires all efforts to preserve it. Hypothesis evaluation of the abiotic variable–probability of occurrence relationships we did by using the generalized additive model with a Bayesian approach. Small climatic intervals we observed and response curves were asymmetric in most cases. Variables that showed statistical significance in the analyses were the mean temperature in the warmest month, mean annual temperature and the summer dryness index. The findings contribute to the scientific basis for the conservation and management of the species and could implement specific actions, such as the exact location of sites in the wildlife for assisted propagation trials, to establish or expand natural protected areas based on the altitudinal gradient or suitable climate ranges.

Keywords: Bayesian regression, biodiversity, climatic change, suitable temperature, *Taxus globosa*

Introduction

One of the species with the most restricted habitat in the world is *Taxus globosa* Schltdl (Taxaceae) (Cope 1998) whose distribution is limited to small sites forming patches, strips or populations with isolated individuals in Mexico, Guatemala, El Salvador and Honduras (Cope 1998, Contreras-Medina and Vega-Luna 2001, Shemluck et al. 2003). Its habitat is in a critical state and its populations have decreased drastically in recent decades (Barrera 1990) and are currently classified as endangered (EN) in the wild (IUCN 2019) as well as in the Mexican Official Standard NOM-059-SEMARNAT-2010 in the category of 'special protection' (SEMARNAT 2010). Besides being a species of high conservation interest, *Taxus globosa* stores paclitaxel (taxol) in its organs, mainly in the bark, twigs and leaves (Soto-Hernández et al. 2000), which is used in cancer treatments including cervical cancer, lung cancer and breast cancer (Saville et al. 1995, Soto-Hernández et al. 2000, Barquero 2007).

NORDIC SOCIETY OIKOS

www.nordicjbotany.org

© 2021 The Authors. Nordic Journal of Botany published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Studies of the species have focused on its distribution from an overall perspective (Muñoz-Gutiérrez et al. 2019), identifying potential areas of distribution (Contreras-Medina et al. 2011, García-Aranda et al. 2012), on its demography (Zavala-Chávez 2001); on the morphological variation of its seeds (Ramírez-Sánchez et al. 2011) or studies on potential uses (Wheeler and Hehnen 1993, Soto-Hernández et al. 2000), among others. Yet, some topics about this species need to be further developed, including the taxol's biosynthetic process (Ramirez-Estrada et al. 2015); its reproductive cycle and dispersal mechanisms (López-Herrera 2008) as well as its relationship with the abiotic variables that prevail in their natural habitat (López-Upton and García-Martí 2015). For instance, the definition of the optimum temperature and the amount of humidity (rainfall) required at different times of the year, could be useful in designing mechanisms of propagation of the species, including the artificial propagation; it would strengthen the underlying theory on the ecological relationships between *T. globosa* and the environmental variables that define its distribution and abundance.

When studying populations at risk of extinction, it is difficult to have very large samples, since a continuous collection of specimens or conducting censuses is costly and demands extra effort most times (Ortega-Argueta and Contreras-Hernández 2013). Fortunately, different analytical methods have been explored in the constant search for reliable and robust results. Generalized linear models (GLM) (Nelder and Wedderburn 1972) being among the most consistent correlative models for detecting the variables with the highest incidence in the distribution of plants regardless of whether the data included in their structure follow a specific theoretical distribution, such as the normal distribution (Guisan et al. 2002, Ravindra et al. 2019). Considering the limited availability of data on species at risk of extinction (such as *Taxus globosa*), the GLM with a Bayesian approach offers some additional advantages, including the possibility of robust modeling of data from small samples (Uusitalo 2007); but in addition, offers a causal explanation between the predictors and the variable of interest because it is a technique for learning, avoiding the construction of biased models (Heckerman 2008). Since the Bayesian approach bases its calculations on posterior probability and credible intervals (Dunson 2001), it offers intuitive, unbiased and consistent results when examining the relationships between the variables studied (Dunson 2001, Fordyce et al. 2011). Here, Bayesian generalized linear models we applied to: 1) identify the variables (out of 29) that have a high and significant importance on the probability of occurrence of *Taxus globosa* and 2) determine the interval of the highest probability of occurrence of this species as a function of the variables that have a significant importance on it.

Material and methods

Description and geographical distribution of the species studied

Taxus globosa Schltdl. (Taxaceae), is an evergreen a gymnosperm with many hanging branches forming a rounded or extended crown (López-Herrera 2008). From its stem, a light brown scaly bark is detached (Luna-Vega 2003) and can reach a total height of 18 m and a stem diameter (at breast height) of up to 80 cm in adulthood (Luna-Vega 2003, Earle 2020). It grows in sites with well-drained acidic soils, with abundant inorganic matter and commonly cohabits with some broadleaf trees, mainly of the genus *Quercus* and some conifers belonging to the genera *Abies*, *Pinus* and *Picea* (Zavala-Chávez 2001, Luna-Vega 2003, Earle 2020). It has been observed at altitudes between 1000 and 3000 m, and its preference sites are small glens where the average annual temperature ranges between 14.5 and 15.4°C, often with fog, dew and frost (Zavala-Chávez 2001). Geographically, it is distributed along the Sierra Madre Oriental of Mexico and some parts of the Sierra of Oaxaca and Chiapas, as far as Central America, covering some localities of Guatemala, Honduras and El Salvador (Fig. 1). Is currently listed as an endangered species (EN) by the International Union for Conservation of Nature (IUCN) (IUCN 2019) and has been catalogued under the category of Subject to Special Protection (Pr) by the Official Mexican Standard NOM-059-SEMARNAT-2010 (SEMARNAT 2010).

Sampling and data collection

We integrated the database using two subsets of data. The first, corresponds to presence records got from portals that collates digitized collection and survey data worldwide, including the Global Biodiversity Information Facility (GBIF) (<www.gbif.org>), from which 74 useful records we got for the study (after debugging); tropicos database from Missouri botanical garden (<https://tropicos.org>) (with 23 usable records), the Gymnosperm database (<www.conifers.org/>) (with 23 usable records), a web's source of information on conifers; as well as from Sánchez-Presa's work (2012) (3 records), from López-Upton and García-Martí (2015) (74 records) and Pérez-Cruz (2018) (24 records), adding up to a total of 221 georeferenced points of presence (Universal Transverse Mercator coordinates, WGS84 datum). All points were distributed in the Americas including four countries: Mexico, Guatemala, El Salvador and Honduras (Fig. 1). All information obtained from different sources was thoroughly reviewed to eliminate inconsistent or duplicate data.

The pseudo-absence data were obtained from the National Forest Inventory (2013–2016) database provided by the National Forestry Commission of Mexico (CONAFOR), which has a network of main plots called conglomerates, distributed throughout the Mexican territory, established every five kilometers in temperate forests, in high jungles and in sub-medium jungles; every 10 km in deciduous forests; and every 20 km in arid zones (CONAFOR 2009, 2015). To avoid bias when fitting the regression models, the selection of pseudo-absence points was equal to the number of points of presence (221) (VanDerWal et al. 2009). To obtain georeferenced points with contrasting ecological characteristics of *Taxus globosa* habitats, we made two filters. First, following the suggestion of VanDerWal et al. (2009), by a systematic, sites at a distance greater than 200 km from the points

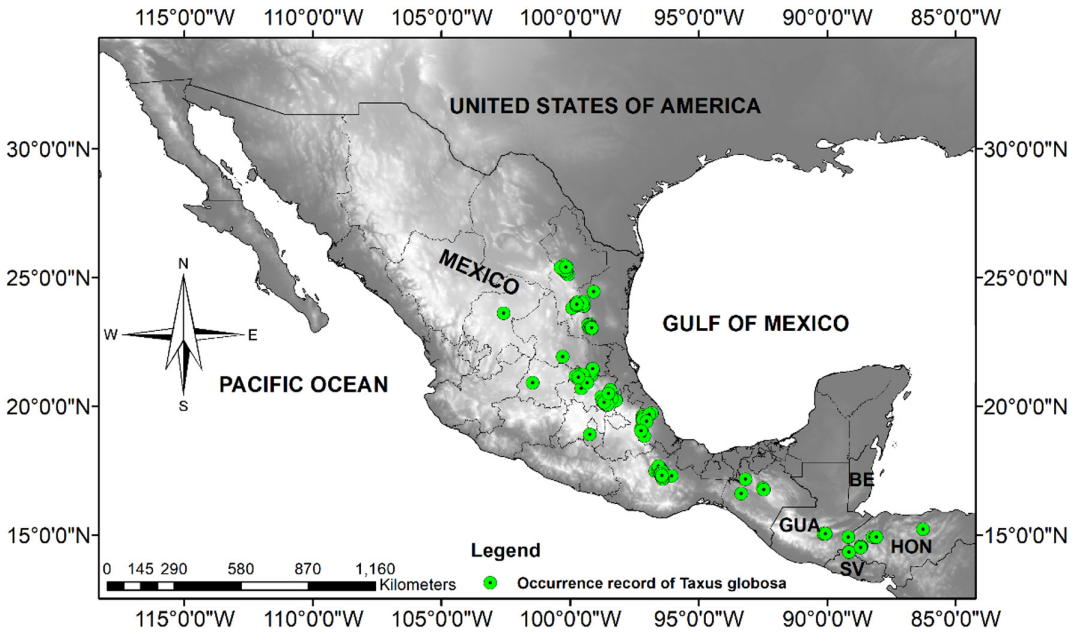


Figure 1. Study area showing the occurrence record of *Taxus globosa*. BE: Belize, GUA: Guatemala; HON: Honduras; SV: El Salvador.

of presence or located outside the potential area projected by Contreras-Medina and Vega-Luna (2001) and García-Aranda et al. (2012) we filtered. These pre-selected points we extracted only 221 by random sampling, using the 'sample' function of the R software. Finally, the two subsets of data were brought together to form a single database with points of presence–absence.

Obtaining the abiotic variables

We used 29 abiotic variables including 23 climatic variables, three soil chemical properties and three physiography variables, registered or estimated for each point (Table 1). We choose these variables as they form part of the spectrum of parameters responsible in the distribution and abundance of plants and other organisms and given its relative ease in their calculation, recording or estimation for the points of interest (occurrence records). The climatic variables we got from monthly climatic surfaces built by Rehfeldt (2006) derived from thin-plate splines of the most important climatic variables for plants (Rehfeldt 2006, Rehfeldt et al. 2006) housed on the Forest Service Rocky Mountain Research Station platform (<http://charcoal.cnre.vt.edu/climate/>). For calculating the variables, this model uses historical data from 1961 to 1990 of over 4000 weather stations distributed in the study area, including Mexico, Guatemala, Belize, Cuba and the southern United States (Rehfeldt et al. 2006, Crookston et al. 2008, Sáenz-Romero et al. 2010). This source we chose because it provides climatic values of the sites where the species has been observed (point estimates of climate) (Rehfeldt 2006).

For each sampling point, we acquired three soil variables including the pH value in H_2O (pH), the organic carbon content (OC) and the cation exchange capacity (CEC) at a depth of 0.30–0.60 m. For this, system soil mapping called SoilGrids (<https://soilgrids.org>) we used whose layers were built up from soil profiles at one kilometer spatial resolution (Hengl et al. 2017). This platform hosts a system, which can estimate some chemical and physical properties of soils from global covariates using interpolation models. Details of the collection of soil samples, contributing institutions and research centers, and details of the models used can be consulted at Hengl et al. (2014, 2017) and Shangguan et al. (2016). In addition, for Mexico's sampling points, the exposure and slope we calculated by a digital elevation models (DEMs) provided by National Institute of Statistics and Geography, Mexico (INEGI) (<www.inegi.org.mx>). This page describes that the DEMs were generated by the LIDAR remote sensing method with a resolution of 15 m and an accuracy of 15–90 cm.

Data analysis

In order to know affectation that each predictor exerts on the probability of occurrence of the species studied, the generalized linear model with Bayesian approach (BGLM) we used, because the full range of possible inferential solutions is retrieved, rather than point estimation and a single confidence interval since its calculations are based on posterior probability and credible intervals (Bedrick et al. 1997, Gelman et al. 2008, 2016, Fordyce et al. 2011).

Steps followed in the data analysis:

1. We grouped the 29 variables into four groups (Table 1): (a) temperature, (b) precipitation, (c) physiography and (d) soil variables. In order to get standardized

Table 1. Descriptive statistics of abiotic variables used in the study.

Variables	Units	Min	Max	Mean	SD	Skew	Category
ADI: annual dryness index, sqrt (DD5)/MAP	–	0.45	10.97	3.54	2.05	0.65	PREC
ASP: aspect of the terrain (azimuth in degrees from north)	°	8	360	213	122.5	−0.32	PHYS
CEC: cation exchange capacity of soil at depth 0.30–0.60 m	cmolc kg^{-1}	17	40	25.23	3.15	0.7	PHYS
D100: julian date the sum of degree-days > 5°C reaches 100	degree-days	7	63	16.1	7.95	3.15	TEMP
DD0: degree-days < 0°C (based on mean monthly temperature)	degree-days	0	49	0.65	4.84	8.92	TEMP
DD5: degree-days > 5°C (based on mean monthly temperature)	degree-days	785	6724	3649	1189	0.05	TEMP
ELEV: elevation above sea level	m	352.6	3946	2207	668.6	−0.33	PHYS
FDAY: julian date of the first freezing date of autumn	day	264	364	341.6	34.06	−1.06	TEMP
FFP: length of the frost-free period	days	115	365	306.3	66.91	−0.95	TEMP
GSDD5: degree-days > 5°C accumulating within the frost-free period	degree-days	244	6565	3354	1345	−0.09	TEMP
GSP: growing season precipitation (April–September)	mm	323	1935	990.1	403	0.26	PREC
MAP: mean annual precipitation	mm	407	2744	1303	572.9	0.36	PREC
MAT: mean annual temperature	°C	6.1	23.6	15.01	3.43	−0.03	TEMP
MMAX: mean maximum temperature in the warmest month	°C	12.5	34.2	24.08	4.54	−0.05	TEMP
MMIN: mean minimum temperature in the coldest month	°C	0.4	16.7	5.94	2.79	0.69	TEMP
MMINDD0: degree-days < 0°C (based on mean minimum monthly temperature)	degree-days	0	410	34.4	60.11	4.04	TEMP
MTCM: mean temperature in the coldest month	°C	4.5	20.6	12.09	2.72	0.01	TEMP
MTWM: mean temperature in the warmest month	°C	8	28	17.57	3.96	0.22	TEMP
PH: soil pH value in H_2O at depth 0.30–0.60 m	–	47.5	79.5	58.51	7.76	0.86	SOIL
PRATIO: ratio of summer precipitation to total precipitation (GSP/MAP)	–	0.57	0.88	0.77	0.07	−0.68	PREC
SDAY: julian date of the last freezing date of spring	day	0	149	29.73	29.7	1.36	TEMP
SDI: summer dryness index (GSDD5/GSP)	–	0.18	13.22	4.15	2.43	0.63	PREC
SLO: average slope of the terrain	°	0	51.79	21.88	12.91	0.3	PREC
SMRP: summer precipitation (Jul + Aug)	mm	133	872	389.7	180.3	0.5	PREC
SMRPB: summer precipitation balance: (Jul + Aug + Sep)/(Apr + May + Jun)	–	1.16	2.76	1.86	0.28	0.09	PREC
SMRSPRPB: summer/spring precipitation balance: (Jul + Aug)/(Apr + May)	–	1.21	6.98	2.96	1.18	0.7	PREC
SOC: soil organic carbon content (fine earth fraction) at depth 0.30–0.60 m	g kg^{-1}	4	49	17.4	7.27	1.35	SOIL
SPRP: spring precipitation (Apr + May)	mm	32	261	134.2	46.56	0.9	PREC
WINP: winter precipitation	mm	30	727	163.7	104.9	1.32	PREC

SD: standard deviation, MIN: minimum, MAX: maximum, Skew: kewness value Fisher–Pearson asymmetry coefficient, TEMP: temperature variables, PREC: precipitation variables, PHYS: physiography variables and SOIL: soil variables.

regression coefficients, we scaled each variable, subtracting each value to its average and dividing by its standard deviation.

2. Through the stepwise algorithm, BGLMs we adjusted the BGLM where we eliminated the variables of each group through two procedures: 'forwards' and 'backwards', using as selection criteria the Akaike information criterion (AIC) (Schwarz 1978) following Zhang's suggestion (2016a), filtering only those that appeared in both selection procedures.
3. Using the filtered variables from each group in the previous step, the LMBs were re-adjusted. The performance of each model we tested using the following measures of goodness of fit: 1) the residual deviance (Deviance) which measures the percentage of the variability captured by the model when including each predictor (Guisan and Zimmermann 2000); 2) a pseudo R squared adjusted (McFadden-R^2adj) whose the calculation is based on the logarithmic probability kernels, for both the univariate and the full model (McFadden 1977); 3) the probability value associated with the χ^2 test when doing an analysis of deviance and, 4) the residuals behavior in a quantile–quantile plot (Q–Q plot) (Augustin et al. 2012, Fabozzi et al. 2014, Zhang 2016b). For reducing the probability of making the type I error (false positives) in multiple hypotheses, a Bonferroni correction we made by dividing the significance level (0.05) by the total number of hypotheses (Hochberg 1988).
4. For variables that showed high and significant effect ($p < 0.05$) in the last run of the models, the probability density

curves we constructed. For this, the Gaussian mixture model (GMM) with two components we carried out using R's *mclust* package (Fraley et al. 2012, 2020) and the six-parameter Weibull models (W6p) that we ran in the *cardidates* package of R (Petzoldt et al. 2018, Rolinski et al. 2018). Both models we selected are flexible and accurately describe bioclimatic data for tree and shrub species with unimodal or multimodal trends (Martínez-Antúnez et al. 2015, Antúnez et al. 2017, 2018). These models we evaluated with the Kolmogorov–Smirnov test, observing the value of D value, which measures the maximum absolute differences between the accumulated empirical data and the theoretical distribution of the models (Young 1977). The one with the smallest D value was the most accurate.

5. Empirical data were used to find a 95% confidence interval based on a Bayesian analysis to a posterior distribution (HDR) (Wright 1986, Dixon and Simon 1991, Hyndman 1996), which is called here as a partial optimum interval. We also calculated it the value of the variable where the tangent line slope of the density curve is equal to zero we calculated; i.e. the maximum of the probability density function, referred here as the point of maximum probability (PMP).

The cross validation of the BGLM models we made with 80% of the data for the model training and 20% for the test, which we did in the R *boot* library using *cv.glm* function (Canty and Ripley 2019).

All analyses, including exploratory analysis (correlation analysis, cleaning and filtering), BGLMs, calculation of the maximum of the probability and intervals we executed in the free statistical application R (<www.r-project.org>).

Results

Variables with a major effect on *Taxus globosa*

We observed ten significant variables in the final Bayesian generalized linear models, suggesting that, as the values of the variables change, there is evidence that the probability of occurrence of *Taxus globosa* responds significantly. The variables that explained the highest percentage of variability in the data were the mean temperature in the warmest month (81%) and elevation above sea level (80%) (Table 2). The precipitation variables with the greatest contribution were the summer dryness index, the aridity index, the ratio of summer precipitation to total precipitation and the average record of winter precipitation (mm) explaining together 91% of the data variability while the only soil variable with significant evidence (organic carbon content) explained 2% of the data variability (Table 2).

Six of the ten significant variables in the latest model showed a negative relationship with the probability of occurrence and four showed a positive relationship. The highest coefficients found in SDI (−327.90) and ADI (186.62), although the errors associated with these same variables also increased significantly (Table 2). In examining the multiple models in quantile–quantile plots, the temperature, precipitation and physiography models showed a better fit compared to the soil model because we observed their residuals within the limits of the 95% confidential band (Fig. 2).

Highest probability of presence and optimum intervals

In the evaluation of density curves with the Kolmogorov–Smirnov test, we found lower D values in seven cases with GMM and three with the Weibull model (Table 3). However, when comparing both means of the points of the maximum probability (PMP), there was no significant difference according to the Kruskal–Wallis test, by showing a probability value (p-value) equal to 0.44. Table 3 shows the values where the maximum probability of occurrence occurs, according to the probabilistic models, as well as the relative optimum interval based on the highest density region ($1-\alpha = 95\%$). For example, considering the mean temperature in the warmest month, we observed the highest probability region between 12.40 and 26.30°C and the maximum at 16.10°C. Additional adjustment parameters and indicators for the

Table 2. Estimated parameters and adjustment indicators of the Bayesian generalized linear models.

Variables	Str.Coeff.	SE	p-value	Deviance	AIC	p-val.χ^2	McFadden-R^2adj. Uni.	McFadden-R^2adj. Mul.
MTWM	−13.41	2.48	< 0.0001***	105.26	111.26	< 0.0001***	0.81	0.89
MAT	16.33	3.97	< 0.0001***	75.60	81.60	0.080.	0.75	
MTCM	−9.37	1.97	< 0.0001***	83.21	89.21	< 0.0001***	0.51	
SDI	−327.90	53.15	< 0.0001***	81.04	99.04	< 0.0001***	0.59	0.91
ADI	186.62	31.14	< 0.0001***	63.60	81.60	< 0.0001***	0.48	
PRATIO	−16.11	2.57	< 0.0001***	45.56	63.56	< 0.0001***	0.16	
WINP	−5.98	1.01	< 0.0001***	19.56	37.56	< 0.0001***	0.01	
SLOP	0.67	0.35	0.05*	88.22	92.22	< 0.0001***	0.46	0.81
ELEV	4.20	0.60	< 0.0001***	241.18	245.18	< 0.0001***	0.80	
SOC	−0.35	0.12	0.002**	434.69	438.69	0.002**	0.02	

Str.Coeff: standardized coefficients, SE: standard error of coefficients, p-value: probability value, Deviance: value that measures the percentage of variability of the data when the predictor is included in the model, AIC: Akaike information criterion due to the predictor, p-val.χ^2: probability value of the χ^2 test, McFadden-R^2adj.Uni: McFadden's pseudo-R squared when including a single variable, McFadden-R^2adj.Mul: McFadden's pseudo-R squared when including the variables of each group, MTWM: mean temperature in the warmest month, MAT: mean annual temperature, MTCM: mean temperature in the coldest month, SDI: summer dryness index, ADI: annual dryness index, PRATIO: ratio of summer precipitation to total precipitation, WINP: winter precipitation, SLOP: average slope of the terrain, ELEV: elevation above sea level, SOC: soil organic carbon content.

two probabilistic models are provided in the Supplementary information.

The density curves also revealed the type of distribution of the data in the face of unitary variation of the abiotic variables, with asymmetric distributions observed in most cases. Figure 3a–b show the optimum points computed by the two probabilistic models (GMM and W6p), the limits of the relative optimum intervals and the distributions of the empirical data. We observed a symmetric and uniform distribution of the unit change in average temperature during the warmest month, but a totally asymmetrical trend in winter precipitation (Fig. 3b).

Representation of probability regions

Using the ten most important variables, we plotted probabilistic regions and, in all cases, asymmetric regions we observed. For example, when combining the mean temperature in the warmest month and the winter precipitation (temperature and precipitation variables with signs of significant incidence on the species), there is a second peak with a smaller-scale that represents a high presence of the species studied (Fig. 4a–b). A combination of three variables showed more clearly the asymmetry of the resulting probabilistic region, even combining variables that separately showed relatively symmetrical trends such as the three temperature variables (Fig. 4c–d).

Discussion

Response to climatic variables

This study defined suitable climatic, physiography and soil values, based on the probability of maximum occurrence for *Taxus globosa*, an endangered species (IUCN 2019) whose populations have declined rapidly in recent decades (Barrera 1990, SEMARNAT 2010). The study also attempted to measure the degree of incidence of 29 abiotic variables and found that, of the total, ten had a significant effect ($p < 0.05$) on the probability of occurrence of the species studied (Table 2). The results shown here represent an important advance in the knowledge of the bioclimatic space of *Taxus globosa*, and constitute a small step in the understanding of its thermal tolerance, its basal thermotolerance and the general resilience of the species.

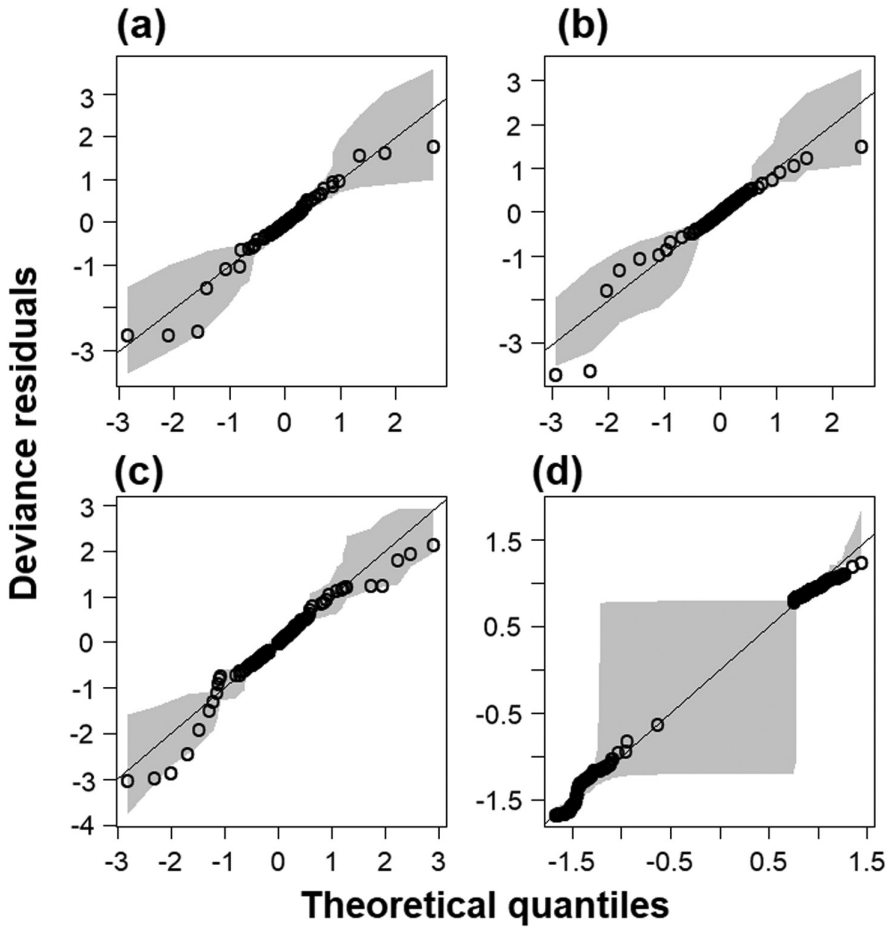


Figure 2. The standardized residuals on a quantile–quantile plot by groups of variables. (a) Temperature variables, (b) precipitation variables, (c) physiographic variables and (d) soil variables.

Table 3. Bioclimatic values where the probability of occurrence of *Taxus globosa* is the maximum and its partial optimum intervals (1–∝=95%) according to the Gaussian mixture model (GMM) and the six-parameter Weibull model (W6p). Maximum probability of occurrence values with the smallest D values of the two models are highlighted in gray.

Variables	Lower limit	Maximum probability of occurrence		Upper limit	Interval	D of GMM	D of W6p
		GMM	W6P				
MTWM: mean temperature in the warmest month (°C)	12.40	16.10	15.51	26.30	13.90	0.26	0.58
MAT: mean annual temperature (°C)	10.00	15.00	14.66	21.80	11.80	0.24	0.06
MTCM: mean temperature in the coldest month (°C)	8.10	12.10	11.74	17.70	9.60	0.26	0.12
SDI: summer dryness index	0.18	2.49	1.23	7.92	7.74	0.35	0.36
ADI: annual dryness index	0.45	1.74	1.22	6.59	6.14	0.35	0.09
PRATIO: ratio of summer precipitation to total precipitation	0.71	0.77	1.00	0.85	0.14	0.41	0.81
WINP: winter precipitation (mm)	30.00	85.00	66.01	334.00	304.00	0.48	0.99
SLOP: average slope of the terrain (°)	1.96	16.96	18.48	42.65	40.69	0.22	0.80
ELEV: elevation above sea level (m)	642.74	2500.00	2949.85	3101.11	2458.37	0.22	0.97
SOC: soil organic carbon content	7.00	14.50	14.33	32.50	25.50	0.40	0.76

The letter D in the last two columns represents the Kolmogorov–Smirnov statistic that measures the maximum distance between the empirical cumulative distribution function of the data and the cumulative distribution function of the Gaussian mixture model (GMM) and the six-parameter Weibull model (W6p).

The BGLMs, suggest a high sensitivity of the probability of *Taxus globosa* registration against the unit change of extreme temperatures, especially, warm and cold months. An average response is 13.42 units for each degree of variation of the mean temperature in the warmest month (May), and a response of 9.37 units for each degree of variation of the mean temperature in the coldest month (January), in both cases, there seems to be a negative relationship (Table 2). These results are in consonance with Moir (1999), who observed a significant effect of summer temperature on the growth of *Taxus baccata*. The same author reports an adverse effect of the low temperatures of January and February on this species. However, low temperatures also encourage the rooting of yews in the early growth stage (Muñoz-Gutiérrez et al. 2009). The case of the warmest month has not only been referred to as an important variable in the growth and development of forest species (Schweingruber 1986), but also as one that defines the distribution and richness of a good part of the plant communities (Gwitira et al. 2014).

Moisture appears to play an important role in yew trees, including an influence on seed morphology and winter dormancy (Ramírez-Sánchez et al. 2011) or encourage seed growth (Moir 1999). Other variables, such as the dryness index and summer humidity, also affect yews; in special, the ratio of summer precipitation to total precipitation

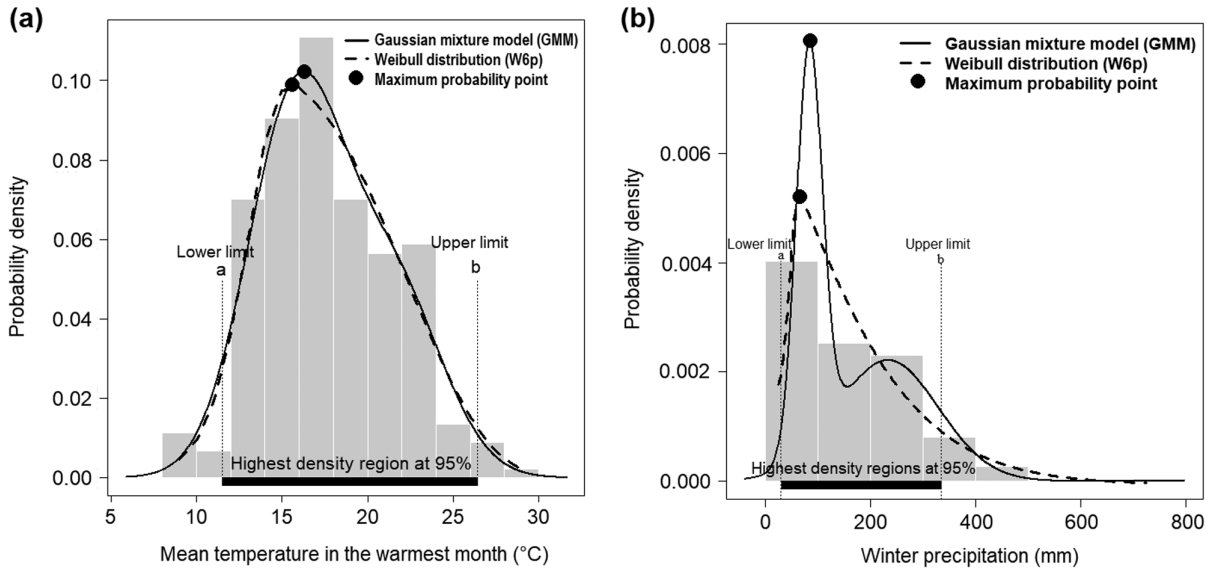


Figure 3. Density curves and graphical representation of the relative optimum interval of *Taxus globosa* as a function of (a) mean temperature in the warmest month and (b) winter precipitation.

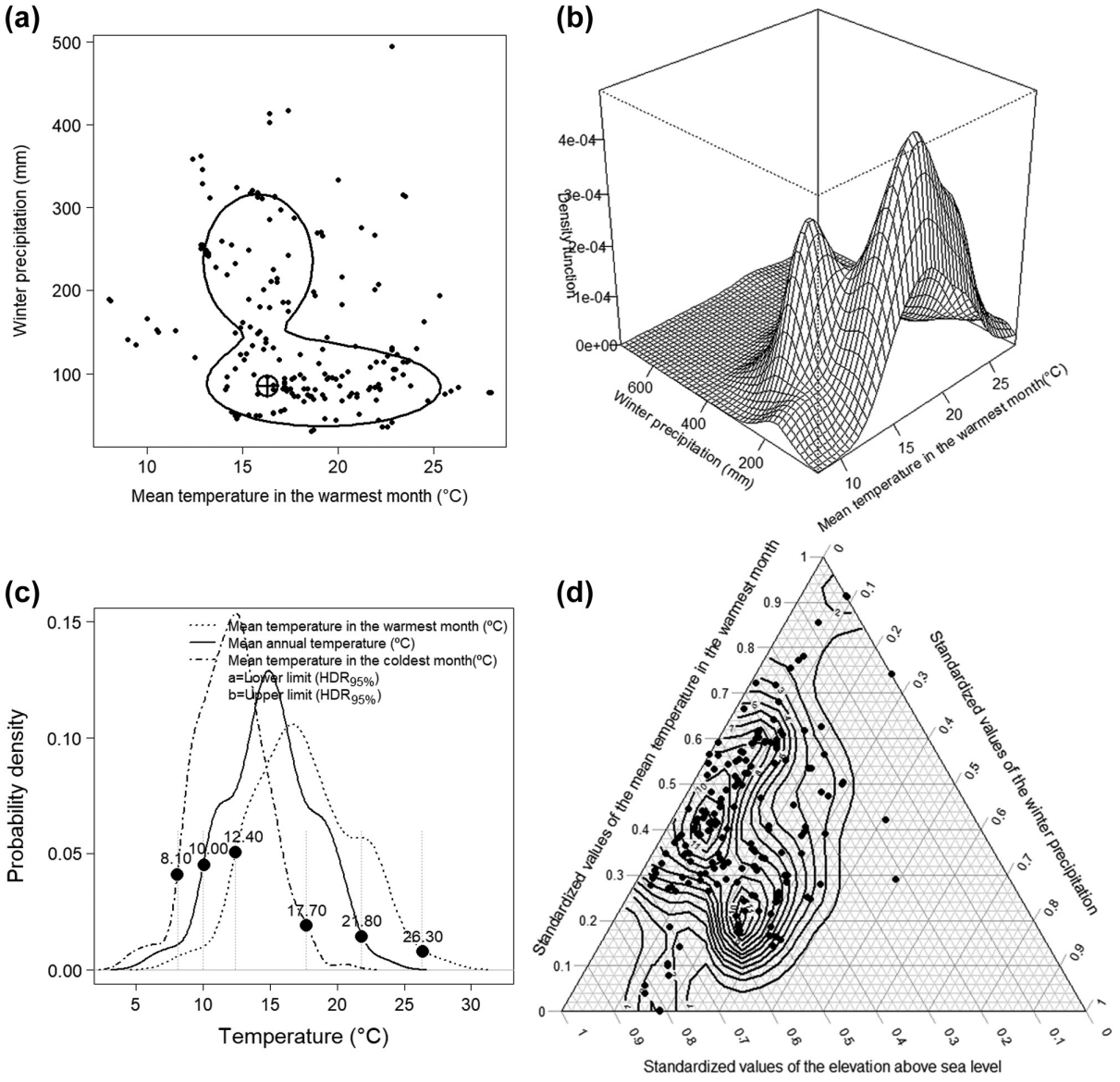


Figure 4. (a) Bidimentional and (b) perspective representation of the optimum probability region for *Taxus globosa* defined by two climatic variables. Relative optimum intervals (c) without interaction of three temperature variables and (d) representation of probability regions by combining the three standardized temperature variables. In (a), the cross mark represents the point where the maximum probabilities of the two interacting variables converge.

(PRATIO) and summer dryness index (SDI = GSDD5/GSP) (Table 3). The latter showed the greatest effect, explaining 59% of the change in the data. Precipitation during specific periods could be related to regulating of the temperature of the yew microhabitat, since the sites where *Taxus globosa* commonly grows are humid glens or shady places (Ramírez-Sánchez et al. 2011), and some populations, even have creeks or waterbodies (Pérez-Cruz 2018).

Response to physiographic and soil variables

In this study, the elevation above sea level explains 80 percent of the variability in the data, becoming one of the most important factors for the studied species. *T. globosa* is abundant between 643 and 3101 m above sea level, with an absolute maximum at 2500 m. These values are partly similar to the altitudinal range reported in previous studies, which fluctuate between 1000 and 3000 m (López-Herrera 2008, Ramírez-Sánchez et al. 2011, García-Aranda et al. 2012). The estimated maximum probability point (2500), varied by 100 m from the optimum reported by Pérez-Cruz (2018) which was 2600 m (Table 3).

Another physiographic variable that showed significant evidence was the slope of the terrain, with a mean response of 0.67 units for each degree of variation in the mean slope of the sites, with a greater proportion of *Taxus globosa* observed between 1.96 and 42.65° and a maximum probability at

the slope of 16.96°. The importance of the slope lies in its close link to nutrient retention and soil moisture (Bridge and Johnson 2000). Findings show that the slope is positive and significantly associated with the probability of *Taxus globosa* occurrence and explains 45% of the data variance (Table 2). Contrary to what has been observed in Europe, where geographical exposure is significant but the slope of the soil is not crucial (Tittensor 1980, Thomas and Polwart 2003).

In relation to the soil variables, soil organic carbon content (fine earth fraction) was the only one showing significant evidence, with a response of 0.35 units in the probability of occurrence for each unit of change of this variable. The maximum probability of occurrence was 14.50 and the limits of the relative optimum interval were 7.00 (lower) and 25.50 (upper) (Table 3). We found no studies related to organic carbon content; however, the soils described for yews in the literature are moist or very moist (Moir 1999) with a neutral to alkaline pH (Shemluck et al. 2003). The ideal pH values reported ranges from 4.1 to 6.3 (López-Herrera 2008, Pérez-Cruz 2018), while the organic matter, content varies among each population, ranging from 3.8% (Pérez-Cruz 2018) to 26.1% (López-Herrera 2008).

Response curves and environmental suitability intervals

Most response curves and probabilistic areas reported here follow asymmetric distributions (Fig. 3b, 4a–b, d) which could reflect the species' discontinuous distribution in physical and environmental space. These results are similar to those of Vetaas (2000) and Antúnez et al. (2017) when investigating the response of some tree and shrub species to unitary changes in temperature or precipitation at specific times of the year. In some cases, not only asymmetric distributions, but also bimodal or multimodal distributions have been observed (Whittaker 1967, Antúnez et al. 2018). The ideal projected intervals for many of the variables, including slope, temperature and precipitation variables, were smaller than the estimated intervals for other conifers like *Pinus patula* Schl. et Cham or *Pinus pseudostrobus* Lindl. (Antúnez et al. 2018).

Some advantages of the graphical results include the detection of the turning points of the dependent variable by increasing or decreasing the predictor (Fig. 3a–b), whereas the partial optimum interval delimits the plausible abiotic range for the growth and development of the species. It could therefore be said that Fig. 4d delimits the environment-and-niche hyperspace of *T. globosa* based on their three most relevant temperature variables, while Fig. 4a–b, as well as showing the ideal probabilistic region, can also detect a second vertex where the occurrence of the species is high, but not so high as the point of maximum probability. In general, probability diagrams and response curves are important because they suggest: 1) the plausible range of climatic values where the occurrence of the species under study is maximized; 2) the partial hyperspace delimited by the variables when combined; and, 3) the discontinuities in topography, the possible variations of parental materials in soil (difficult to measure), and the possible changes due to habitat disturbance.

Conclusions and suggestions

One of the most significant ecological challenges currently is the gradual and accelerated loss of biological diversity with increasing rates of extinction (Kuussaari et al. 2009, Davies et al. 2011). Knowing the response of a genetic resource at risk of extinction to the variations of a variable, as well as its range of maximum presence, are important to understand its capacity for resilience. Based on the findings, the most significant temperature variables that enhance or inhibit *Taxus globosa* are the maximum temperature of the hottest month and the temperature of the coldest month. The most relevant precipitation variables, which in most cases seem to have an adverse impact, are the precipitation during the summer and winter. On the other hand, the relative optimum intervals found in this work represent an approximation to the environment-and-niche hyperspace of the studied species which could represent an advance in the delimitation of their real bioclimatic space.

These results, besides contributing to the knowledge of *T. globosa*' ecology, these could also be useful in the implementation of specific conservation actions, e.g. the optimal points and intervals (Table 3) would facilitate the location of potential new sites for a possible assisted propagation of the species in the wildlife. The delimitation of the altitudinal gradient would be useful to establish or expand natural areas of protection or conservation, as well as to delimit biological corridors taking into account species that co-exist or associate with the studied species. Could also be used in actions to reduce the impacts of anthropogenic activity, including environmental education or social works in the communities near *T. globosa*'s distribution areas.

Acknowledgements – Open Access funding enabled and organized by Projekt DEAL.
Funding – National Council of Science and Technology (CONACYT).
Conflict of interests – There is no conflict of interest of any nature.

References

Antúnez, P. et al. 2017. The role of climatic variables for estimating probability of abundance of tree species. – Pol. J. Ecol. 65: 324–338.

Antúnez, P. et al. 2018. Density functions: an application for delimiting optimal intervals of climate and physiography for forest species. – Agrociencia 52: 1031–1042.

Augustin, N. H. et al. 2012. On quantile quantile plots for generalized linear models. – Comput. Stat. Data An. 56: 2404–2409.

Barquero, A. 2007. Medicinal plants: past, present and future [Plantas sanadoras: pasado, presente y futuro]. – <https://ri.conicet.gov.ar/handle/11336/98477>, accessed 16 Jan 2018.

Barrera, G. M. 1990. Rare, threatened and endangered vascular plants in Tamaulipas [Plantas vasculares raras, amenazadas y en peligro de extinción en Tamaulipas]. – <http://ecologia.uat.edu.mx/paginaiea/biotam/v2n2/art7.html>, accessed 23 Apr 2020.
Bedrick, E. J. et al. 1997. Bayesian binomial regression: predicting survival at a trauma center. – Am. Stat. 51: 211–218.
Bridge, S. R. J. and Johnson, E. A. 2000 Geomorphic principles of terrain organization and vegetation gradients. – J. Veg. Sci. 11: 57–70.
Canty, A. and Ripley, B. 2019. Package 'boot'. – <ftp://ftp.math.ethz.ch/sfs/R-CRAN/web/packages/boot/boot.pdf>, accessed 18 Nov 2019.
CONAFOR (National Forestry Commission) 2009. National forest and soil inventory. Manual and procedures for field sampling. – <https://snigf.cnf.gob.mx/wp-content/uploads/Documentos%20metodologicos/2004-2009/Manual%20y%20procedimientos%20muestreo.pdf>, accessed 17 Jan 2018.
CONAFOR (National Forestry Commission) 2015. National forest and soil inventory. Manual and procedures for field sampling, ver. 158 [Inventario Nacional Forestal y de Suelos-Procedimientos de muestreos Versión 158]. – <www.cnf.gob.mx:8090/snif/portal/component/phocadownload/category/267-2015?download=1228:anexo-procedimientos-de-muestreo-infys-2015>, accessed 17 Jan 2018.
Contreras-Medina, R. and Vega-Luna, I. 2001. Presence of *Taxus globosa* Schltdl (Taxaceae) in the State of Chiapas, Mexico [Presencia de *Taxus globosa* Schltdl (Taxaceae) en el Estado de Chiapas, México]. – Polibotanica 12: 51–55.
Contreras-Medina, R. et al. 2011. The representation of the yew (*Taxus globosa* Schltdl) in the protected natural areas of Mesoamerica [Representatividad del tejo (*Taxus globosa* Schltdl) en las áreas naturales protegidas de Mesoamérica]. – Spanish J. Rural Dev. 3: 51–60.
Cope, E. A. 1998. Taxaceae: the genera and cultivated species. – Bot. Rev. 64: 291–322.
Crookston, N. L. et al. 2008. FVS and global warming: a prospectus for future development. – In: Havis, R. N. and Crookston, N. L. (eds), Proc. 3rd forest vegetation simulator conference. Dept of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA, pp. 7–16.
Davies, T. J. et al. 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. – PLoS Biol. 9: e1000620.
Dixon, D. O. and Simon, R. 1991. Bayesian subset analysis. – Biometrics 1991: 871–881.
Dunson, D. B. 2001. Commentary: practical advantages of Bayesian analysis of epidemiologic data. – Am. J. Epidemiol. 153: 1222–1226.
Earle, C. J. 2020. Gymnosperm database, (copyright 2020). – <https://conifers.org>, accessed 10 Apr 2020.
Fabozzi, F. J. et al. 2014. Appendix E: Model selection criterion: AIC and BIC the basics of financial econometrics. – Wiley, Hoboken, NJ, USA, pp. 399–403, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118856406.app5>, accessed 23 Nov 2019.
Fordyce, J. A. et al. 2011. A hierarchical Bayesian approach to ecological count data: a flexible tool for ecologists. – PLoS One 6: e26785.
Fraley, C. et al. 2012. Package 'mclust'. – <http://ftp.uwsg.indiana.edu/CRAN/web/packages/mclust/mclust.pdf>, accessed 20 Nov 2019.
Fraley, C. et al. 2020. Package 'mclust' ver 5.4.6. – <https://mclust-org.github.io/mclust/>, accessed 25 Apr 2020.
García-Aranda, M. A. et al. 2012. Distribución actual y potencial de *Taxus globosa* (Taxaceae) en México. – J. Bot. Res. Inst. Tex. 2012: 587–598.
Gelman, A. et al. 2008. A weakly informative default prior distribution for logistic and other regression models. – Ann. Appl. Stat. 2: 1360–1383.
Gelman, A. et al. 2016. Package 'arm'. – <https://cran.r-project.org/web/packages/arm/arm.pdf>, accessed 02 Mar 2020.
Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.
Guisan, A. et al. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. – Ecol. Model. 157: 89–100.
Gwitira, I. et al. 2014. Precipitation of the warmest quarter and temperature of the warmest month are key to understanding the effect of climate change on plant species diversity in Southern. – Afr. J. Ecol. 52: 209–216.
Heckerman, D. 2008. A tutorial on learning with Bayesian networks. – In: Innovations in Bayesian networks. Springer, pp 33–82, <https://bi.snu.ac.kr/Courses/4ai15s/slides/Bayesian_Networks.pdf>, accessed 07 May 2020.
Hengl, T. et al. 2014. SoilGrids1km – global soil information based on automated mapping. – PLoS One 9: e105992.
Hengl, T. et al. 2017. SoilGrids250m: global gridded soil information based on machine learning. – PLoS One 12: e0169748.
Hochberg, Y. 1988. Sharper Bonferroni procedure for multiple tests of significance. Biometrika 75: 800–802.
Hyndman, R. J. 1996. Computing and graphing highest density regions. – Am. Stat. 50: 120–126.
IUCN (International Union for Conservation of Nature) 2019. The IUCN Red List of threatened species ver. 2019-2. – <www.iucnredlist.org>, accessed 15 Sep 2019.
Kuussaari, M. et al. 2009. Extinction debt: a challenge for biodiversity conservation. – Trends Ecol. Evol. 24: 564–571.
López-Herrera, M. 2008. *Taxus globosa* Schltdl: a medicinal species in El Chico National Park, Hidalgo [*Taxus globosa* Schltdl una especie medicinal en el parque nacional el Chico, Hidalgo]. – <www.uaeh.edu.mx/investigacion/icbi/LI_EcofisPlanSupCulTej/maritza_lopez/taxus.pdf>, accessed 06 Jun 2020.
López-Upton, J. and García-Martí, X. 2015. *Taxus globosa* Schltdl (Taxaceae) distribution and diagnosis of an endangered yew. – Earth 4: 80–88.
Luna-Vega, M. I. 2003. *Taxus globosa*: Taxa of the mesophilic mountain forest of the Sierra Madre Oriental included in the Norma Oficial Mexicana [*Taxus globosa*. Taxones del bosque mesófilo de montaña de la Sierra Madre Oriental incluidos en la norma oficial mexicana]. – <www.conabio.gob.mx/conocimiento/ise/fichasnom/Taxusglobosa00.pdf>, accessed 17 Apr 2020.
Martínez-Antúnez, P. et al. 2015. Use of the Weibull function to model maximum probability of abundance of tree species in northwest Mexico. – Ann. For. Sci. 72: 243–251.
McFadden, D. 1977. Quantitative methods for analyzing travel behavior of individuals: some recent developments. – <https://core.ac.uk/download/pdf/6448852.pdf>, accessed 09 May 2020.
Moir, A. K. 1999. The dendrochronological potential of modern yew *Taxus baccata* with special reference to yew from Hampton Court Palace, UK. – New Phytol. 144: 479–488.

Muñoz-Gutiérrez, L. et al. 2009. Effect of cutting age and substrate temperature on rooting of *Taxus globosa* – New For. 38: 187–196.
Muñoz-Gutiérrez, L. et al. 2019. Review on the distribution and conservation of *Taxus globosa* Schltdl. (Taxaceae) in Mexico. – Rev. Mex. Cie. For. 10: 65–84.
Nelder, J. A. and Wedderburn, R. W. 1972. Generalized linear models. – J. R. Stat. Soc. Ser. A 135: 370–384.
Ortega-Argueta, A. and Contreras-Hernández, A. 2013. Proposal for monitoring and evaluation of programs focused on recovering species at risk of extinction [Propuesta de un esquema de seguimiento y evaluación para programas de recuperación de especies en riesgo]. – Gest. Polit. Publica 22: 457–496.
Pérez-Cruz, N. 2018. Characterization of the soil and environmental conditions of three tree species with high conservation value in the Sierra Norte of Oaxaca [Caracterización parcial de condiciones edafológicas y ambientales de tres especies arbóreas de alto valor de conservación en la Sierra Norte de Oaxaca]. – Bachelor thesis, Univ. of the Sierra, Juarez, Ixtlan de Juarez, Oax. Mexico.
Petzoldt, T. et al. 2018. Manual for Package cardidates. – <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.645.1557&rep=rep1&type=pdf>, accessed 23 Nov 2019.
Ramirez-Estrada, K. et al. 2015. Changes in gene transcription and taxane production in elicited cell cultures of *Taxus* x media and *Taxus globosa*. – Phytochemistry 117: 174–184.
Ramírez-Sánchez, S. E. et al. 2011. Study on the morphological variation of the seeds of *Taxus globosa* Schltdl in two Mexican regions [Variación morfológica de semillas de *Taxus globosa* Schltdl provenientes de dos regiones geográficas de México]. – Rev. Fitotec. Mex. 34: 93–99.
Ravindra, K. et al. 2019. Generalized additive models: building evidence of air pollution, climate change and human health. – Environ Int. 132: 104987.
Rehfeldt, G. E. 2006. A spline model of climate for the western United States Gen Tech Rep RMRS-GTR-165. – US Dept of Agriculture, Forest Service, Rocky Mountain Research Station 21, Fort Collins, CO, USA.
Rehfeldt, G. E. et al. 2006. Empirical analyses of plants climate relationships for the western United States. – Int. J. Plant. Sci. 167: 1123–1150.
Rolinski, S. et al. 2018. Package 'cardidates'. – <http://ftp2.pl.vim.org/pub/Linux/software/mirrors/CRAN/web/packages/cardidates/cardidates.pdf>, accessed 23 Nov 2019.
Sáenz-Romero, C. et al. 2010. Spline models of contemporary, 2030, 2060 and 2090 climates for Mexico and their use in understanding climate-change impacts on the vegetation. – Clim. Change 102: 595–623.
Sánchez-Presa, L. A. 2012. Characterizing the habitat of *Taxus globosa* Schltdl (Taxaceae) in the Alto Pixquiac Basin, Veracruz [Caracterización del hábitat de *Taxus globosa* Schltdl (Taxaceae) en la Cuenca del Alto Pixquiac, Veracruz]. – Ms thesis, Inst. de Ecología AC Xalapa Ver. México.
Saville, M. W. et al. 1995. Treatment of HIV-associated Kaposi's sarcoma with paclitaxel. – Lancet 346: 26–28.
Schwarz, G. 1978. Estimating the dimension of a model. – Ann. Stat. 6: 461–464.
Schweingruber, F. H. 1986. Abrupt growth changes in conifers. – IAWA J. 7: 277–283.
SEMARNAT (Secretariat of Environment and Natural Resources) 2010. Official Mexican Standard NOM-059-SEMARNAT-2010. – <http://dof.gob.mx/nota_detalle_popup.php?codigo=5173091>, accessed 17 Apr 2020.
Shangguan, W. et al. 2016. Mapping the global depth to bedrock for land surface modeling. – J. Adv. Model. Earth Syst. 9: 65–88.
Shemluck, M. J. et al. 2003. A preliminary study of the taxae chemistry and natural history of the Mexican yew, *Taxus globosa* Schltdl. – Bol. Soc. Bot. Mex. 72: 119–127.
Soto-Hernández, M. et al. 2000. The potential use of the Mexican yew (*Taxus globosa* Sch) to extract taxol [El tejo mexicano (*Taxus globosa* Sch). Potencial de su aprovechamiento en taxol]. – Ciencia Ergo Sum 7: 277–279.
Thomas, P. A. and Polwart, A. 2003. *Taxus baccata* L. – J. Ecol. 91: 489–524.
Tittensor, R. M. 1980. Ecological history of yew *Taxus baccata* L in southern England. – Biol. Conserv. 17: 243–265.
Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. – Ecol. Model. 203: 312–318.
VanDerWal, J. et al. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? – Ecol. Model. 220: 589–594.
Vetaas, O. R. 2000. Comparing species temperature response curves: population density versus second-hand data. – J. Veg. Sci. 11: 659–666.
Wheeler, N. C. and Hehnen, M. T. 1993. Taxol: a study in technology commercialization. – J. For. 91: 15–18.
Whittaker, R. H. 1967. Gradient analysis of vegetation. – Biol. Rev. 42: 207–264.
Wright, D. E. 1986. A note on the construction of highest posterior density intervals. – J. R. Stat. Soc. C-Appl. 35: 49–53.
Young, I. T. 1977. Proof without prejudice: use of the Kolmogorov–Smirnov test for the analysis of histograms from flow systems and other sources. – J. Histochem. Cytochem. 25: 935–941.
Zavala-Chávez, F. 2001. Demographic analysis and habitat characterization of *Taxus globosa* Schlecht in El Chico, National Park, Hidalgo, Mexico [Análisis demográfico preliminar de *Taxus globosa* Schlecht en el Parque Nacional El Chico, Hidalgo, México I: Población de adultos y algunas características del hábitat]. – Ciencia Ergo Sum 8: 169–174.
Zhang, Z. 2016a. Variable selection with stepwise and best subset approaches. – Ann. Transl. Med. 4: 136.
Zhang, Z. 2016b. Residuals and regression diagnostics: focusing on logistic regression. – Ann. Transl. Med. 4: 195.