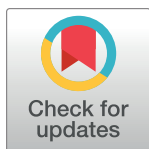


RESEARCH ARTICLE

Dummy regression to predict dry fiber in Agave lechuguilla Torr. in two large-scale bioclimatic regions in Mexico

José Óscar M. López-Díaz¹, Jorge Méndez-González^{2*}, Pablito M. López-Serrano³, Félix de J. Sánchez-Pérez⁴, Fátima M. Méndez-Encina⁵, Rocío Mendieta-Oviedo¹, Librado Sosa-Díaz⁵, Andrés Flores⁶, Emily García-Montiel⁷, Víctor H. Cambrón-Sandoval⁸, Alejandro Zermeño-González⁹, José J. Corral Rivas⁷



1 Consultor Forestal Independiente, Saltillo, Coahuila, México, **2** Departamento Forestal, División de Agronomía, Universidad Autónoma Agraria Antonio Narro, Saltillo, Coahuila, México, **3** Instituto de Silvicultura e Industria de la Madera, Facultad de Ciencias Forestales y Ambientales, Universidad Juárez del Estado de Durango, Durango, México, **4** Consultor Estadístico Independiente, Saltillo, Coahuila, México, **5** Estudiante de Postgrado en Ciencias Forestales, Colegio de Postgraduados, Montecillo, Texcoco, Estado de México, México, **6** Cenid-Comef, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Alcaldía Coyoacán, México, **7** Facultad de Ciencias Forestales y Ambientales, Universidad Juárez del Estado de Durango, Durango, México, **8** Facultad de Ciencias Naturales, Universidad Autónoma de Querétaro, Querétaro, México, **9** Departamento de Riego y Drenaje, División de Ingeniería, Universidad Autónoma Agraria Antonio Narro, Saltillo, Coahuila, México

* These authors contributed equally to this work.

‡ FMME, RMO, LSD, AF, EGM, VHCS, AZG, and JJCR also contributed equally to this work.

* jmendezg@hotmail.com

OPEN ACCESS

Citation: López-Díaz JÓM, Méndez-González J, López-Serrano PM, Sánchez-Pérez FdJ, Méndez-Encina FM, Mendieta-Oviedo R, et al. (2022) Dummy regression to predict dry fiber in Agave lechuguilla Torr. in two large-scale bioclimatic regions in Mexico. PLoS ONE 17(9): e0274641. <https://doi.org/10.1371/journal.pone.0274641>

Editor: Lalit Chandra Saikia, National Institute of Technology Silchar, India, INDIA

Received: March 5, 2022

Accepted: September 1, 2022

Published: September 15, 2022

Copyright: © 2022 López-Díaz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: Funding: This research was founded by the Forestry National Commission (CONAFOR) and National Council of Science and Technology (CONACYT), through project number: 2017-4-29267, titled “Best management practices and generation of volume and biomass equations for the main non-timber forest species of economic

Abstract

Agave lechuguilla Torr., of the family Agavaceae, is distributed from southwestern United States to southern Mexico and is one of the most representative species of arid and semi-arid regions. Its fiber is extracted for multiple purposes. The objective of this study was to generate a robust model to predict dry fiber yield (*Dfw*) rapidly, simply, and inexpensively. We used a power model in its linear form and bioclimatic areas as dummy variables. Training, generation (80%) and validation (20%) of the model was performed using machine learning with the package ‘*caret*’ of R. Using canonical correlation analysis (CCA), we evaluated the relationship of *Dwf* to bioclimatic variables. The principal components analysis (PCA) generated two bioclimatic zones, each with different *A. lechuguilla* productivities. We evaluated 499 individuals in four states of Mexico. The crown diameter (*Cd*) of this species adequately predicts its fiber dry weight ($R^2 = 0.6327$; $p < 0.05$). The intercept (β_0), slope [$\ln Cd$ (β_1)], zone [(β_2)] and interaction [$\ln Cd:Zona$ (β_3)] of the dummy model was statistically significant ($p < 0.05$), giving origin to an equation for each bioclimatic zone. The CCA indicates a positive correlation between minimum temperature of the coldest month (Bio 6) and *Dwf* ($r = 0.84$ and $p < 0.05$). In conclusion, because of the decrease in Bio 6 of more than 0.5°C by 2050, the species could be vulnerable to climate change, and *A. lechuguilla* fiber production could be affected gradually in the coming years.

importance in the arid and semi-arid ecosystems of Mexico". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1. Introduction

In Mexico, arid and semiarid regions cover 54% of the territory and are inhabited by approximately 40% of the country's population [1]. Their vegetation is scrub *Crassicaule*, Desert Microphyll, and Desert Rosetophile; the genera *Larrea*, *Agave*, *Dasylium*, *Yuca*, etc., are the most representative [2]. One of the most economically and ecologically important species in arid regions is *Agave lechuguilla* Torr., a succulent plant of the *Agavaceae* family, distributed from southeastern United States to southern Mexico over 11 federal states [3, 4], covering 142 115 km² [5].

This species, known as "lechuguilla", is outstanding for its fiber ("ixtle"), which is extracted from its bud ("cogollo") once the plant has reached 25 cm in height [6]. The fiber is used in the automobile industry, cordage, rugs, and cleaning brushes, among others [7]. *Lechuguilla* fiber can be extracted manually or mechanically [3, 8]. The former method produces better quality and can sell for a price of up to 1.2 US dollars per kg. With this technique, 1.87 kg h⁻¹ is produced, while the mechanical method can produce between 15 and 18 kg h⁻¹ of lower quality fiber [9]. Better quality fiber is obtained from bud ("cogollos") with young leaves, which contain less lignin than leaves from the plant crown [8]. On *lechuguilla* plantations, the bud takes seven to eight months to regenerate (to reach ≥ 25 cm of height), while naturally it takes between 16 and 24 months [10].

The fiber from *A. lechuguilla* is crucial for inhabitants of arid and semiarid regions of the country [11], benefitting nearly 52,000 families [7]. For this reason, it is necessary to quantify fiber existences. This can be done directly by harvesting the plant or indirectly using allometric models. In Mexico there are only three studies on estimating *A. lechuguilla* fiber dry weight: Blando & Baca [12], Pando *et al.* [13] and Velasco *et al.* [3]. All these studies were conducted at a local scale, and some of these models require harvesting the entire plant or part of it [8, 14] to estimate fiber weight. Since the objective of creating a model is to make predictions with new data, it is necessary that the model comply with all the statistical assumptions, and it is especially important to evaluate its predictive capacity [15].

For this reason, it is crucial to generate robust allometric models to estimate *A. lechuguilla* fiber dry weight with non-destructive techniques that are simple, rapid, and efficient, and that solve the problem of scale, considering the productivity areas and determining the relation of fiber dry weight to bioclimatic variables. The objective of this study was to create a model of this type that will serve not only to support government procedures [16] for authorizing programs for use of the fiber but also for decision-making in management and conservation of the species.

2. Materials and methods

2.1. Description of the study area

The study was conducted in the arid and semiarid regions in northeastern Mexico in the states of Chihuahua (Chih), Coahuila (Coah), San Luis Potosí (SLP) and Zacatecas (Zac). *Agave lechuguilla* grows in colluvial, sandy loam, limestone and clay soils [7, 17] in areas where precipitation varies from 150 to 500 mm, at altitudes from 200 to 2400 m [2, 11] and temperatures of 3 to 30°C [18], and even in extreme temperatures of -8 to 44°C. It is known that this plant can withstand droughts and floods [19].

2.2. Bioclimatic zoning of distribution areas of *Agave lechuguilla* Torr.

We obtained the following bioclimatic variables from the website <https://chelsa-climate.org/>: mean annual temperature, Bio 1 (°C), maximum temperature of the hottest month, Bio 5 (°C),

minimum temperature of the coldest month, Bio 6 ($^{\circ}\text{C}$), and mean annual precipitation, Bio 12 (mm) in raster format at a resolution of 1 km^2 [20]. Using ArcMap 10.4.1, 1000 points were distributed randomly over the area of *A. lechuguilla* distribution in the country, for which the values of each bioclimatic variable and of altitude (m) obtained from the Modelo Digital de Elevación Mexicano (<https://www.inegi.org.mx/app/geo2/elevacionesmex/>) were extracted. With this information, a principal components analysis (PCA) was performed in R software [21] using the library ‘FactoShiny’ [22] which requires ‘FactoMiner’ [23], using the variables in standardized form to extract the first three principal components. Then, with the unrotated eigenvalues derived from PC1, an interpolation (Inverse Distance Weighted) was carried out in ArcMap, from which isolines spaced at 2.85 were derived.

2.3. Sampling *Agave lechuguilla* Torr for fiber extraction

Agave lechuguilla sampling was selective only in “ejidos” with gathering permits. therefore, a special permit was not necessary to harvest the individuals in this study.

Plants ≥ 10 cm tall were collected (according to the gatherers, it is not until the plant reaches this height that the bud can be worked traditionally). Using a 3 m tape measure, we measured average crown diameter (Cd , cm) and total height (H , cm) on each selected plant. Prior to felling the plant (to obtain aerial biomass), the bud was extracted using a “cogollera” (a rustic instrument characteristic of the gatherers) to measure its length (Lc) and immediately extract the fiber manually. The fresh fiber was placed in a paper bag properly labeled and taken to the laboratory at the UAAAN to be dried in a Thermo ScientificTM HERAthemTM (Modelo OMH750) oven at 70°C until it had a constant weight, which was obtained with a Torrey[®] (Modelo L-EQ) scale with a capacity of 5 kg and precision of 1 g.

2.4. Generation and validation of the equation for estimating dry fiber weight

To predict dry fiber (Dfw) of individual *A. lechuguilla* plants, the allometric power equation (Eq 1) was tested in its linear form (Eq 2), applying logarithms to correct error variance [15], as used by Wood [24], Zárate *et al.* [25] and Flores *et al.* [26].

$$Y = \beta_0 X^{\beta_1} \varepsilon \quad (1)$$

$$\ln(Y) = \ln\beta_0 + \beta_1 \times \ln(X) + \varepsilon \quad (2)$$

where Y = fiber dry weight (kg), testing crown diameter (Cd , cm), total plant height (H , cm), and the product of the two, $Cd \times H$ (cm $\times$ cm), as predictor variables, $\ln$ = natural logarithm and β_0 and β_1 = coefficients of regression.

Eq 2 was fit using the bioclimatic zone as the dummy variable $D = \begin{cases} 0 & (\text{Zone 1}) \\ 1 & (\text{Zone 2}) \end{cases}$ considering an effect to the intercept (Zone) and the model slope ($X: \text{Zone}$) to determine the existence of a model (for the entire area) or one for each bioclimatic zone, as follows:

$$\ln(Y) = \ln\beta_0 + \beta_1 \times (\ln X) + \beta_2 \times (\text{Zone}) + \beta_3 \times (\ln X : \text{Zone}) + \varepsilon \quad (3)$$

With 80% of all the data (obtained randomly by quantiles) and using the method of ordinary least squares [15], the model was trained using machine learning (ML) with the package R ‘caret’ [27]. With the same library, the model was then generated by bootstrap ($n = 25$). Hypothesis tests ($\alpha = 0.05$) were performed on the coefficients of regression ($H_0 : \beta_i = 0$ vs $H_1 : \beta_i \neq 0, \dots, \beta_{ij}$) of the final model. Statistical tests ($\alpha = 0.05$) were

applied to assure compliance to assumptions of normality [Lilliefors (Kolmogorov-Smirnov)], variance homogeneity (Breusch-Pagan), and independence of the errors (Durbin-Watson). Possible atypical data ($-3 \leq r_i \leq +3$) were examined through studentized residuals r_i . Influence on the coefficients of the model, a) Cook distance, D_i (observation y_i), and on the precision of COVRATIO estimations [15, 28, 29], and b) DFBETAS ($\hat{\beta}_i$), DFFITS (influence of the i^{th} observation on the predicted values), and on the precision of the estimations. The total of the basic hypotheses of the regression model is summarized in the following expression: $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$; that is, random, independent, and identically distributed errors, according to a normal distribution with mean zero and variance σ^2 .

After verifying compliance to all the previous assumptions, validation of the final model was carried out in 'caret' [27] with 20% of the independent data, using four methods: a) Leave One Out Cross-Validation, b) k-fold validation ($k = 10$), c) Repeated K-fold ($k = 10$ & 3 sets) and d) Bootstrap Cross-Validation, calculating the Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

The return of $\ln$ to its original units to predict *A. lechuguilla* fiber dry weight of the resulting equation is not direct since, if the distribution of a $\log Y$ is normal, a $\log X$, result of the antilogarithm $\log Y$, may be not normal and biased since the median is obtained and not the mean [24]. This would have to be corrected by applying to the final model a correction factor (calculated here), which is given by: $CF = e^{(SEE^2/2)}$, where SEE is the estimation error of the regression [30].

2.5. Canonical correlation analysis

With the aim of identifying the dependence of *A. lechuguilla* dry fiber with environmental variables, as well as the interdependence between factors (matrices X, Y): [Cd, H and Dfw], [Bio 1, Bio 5, Bio 6, Bio 12 and altitude], a Canonical Correlation Analysis was performed using the package R 'CCA' [31] of R statistical software [21]. The matrices X, Y were generated with the average of the variable per municipality (S1 Table).

3. Results and discussion

3.1. Bioclimatic zoning

The results of the PCA showed that components 1 (64.17%) and 2 (25.89%) explain 90.07% of the variance. The contribution of each variable in component 1 was 29.98% (Bio 1), 27.61% (altitude), 21.80% (Bio 6), 17.95% (Bio 5), 2.66% (Bio 12) (Fig 1A). The unrotated scores (eigenvalues = 0.25) gave origin to two bioclimatic zones divided almost longitudinally: Zone 1: Chih, Zac, and the northern part of SLP ($n = 200$); Zone 2: Coah, southeastern SLP and northeastern Zac ($n = 199$) (Fig 1B).

Using the same variables, plus slope, aspect and records of the presence of the species, some authors [32, 33] have generated zones of productive potential in this region of the country, estimating between 9 and 5 million ha that are ideal for *A. lechuguilla* plantations, mainly in zone 2 of our study (Fig 1B). Flores *et al.* [26] constructed models to estimate aboveground biomass of this species, considering the state boundaries (Fig 1) as the dummy variable (or zone) and found differences in biomass among the zones.

When we examined the means of the dendrometric (Cd, H and Dfw) and bioclimatic (Bio 1, Bio 5, Bio 6, Bio 12 and altitude) variables through dummy-regression models equivalent to Student "t" tests) in logarithmic form for compliance to variance normality and homogeneity: $\ln(Y) = \beta_0 + \beta_1 \times \text{Zone}$, we found that there were statistically significant differences among zones ($p \leq 0.05$). In Zone 2, *A. lechuguilla* height, crown diameter, and quantity of fiber were higher

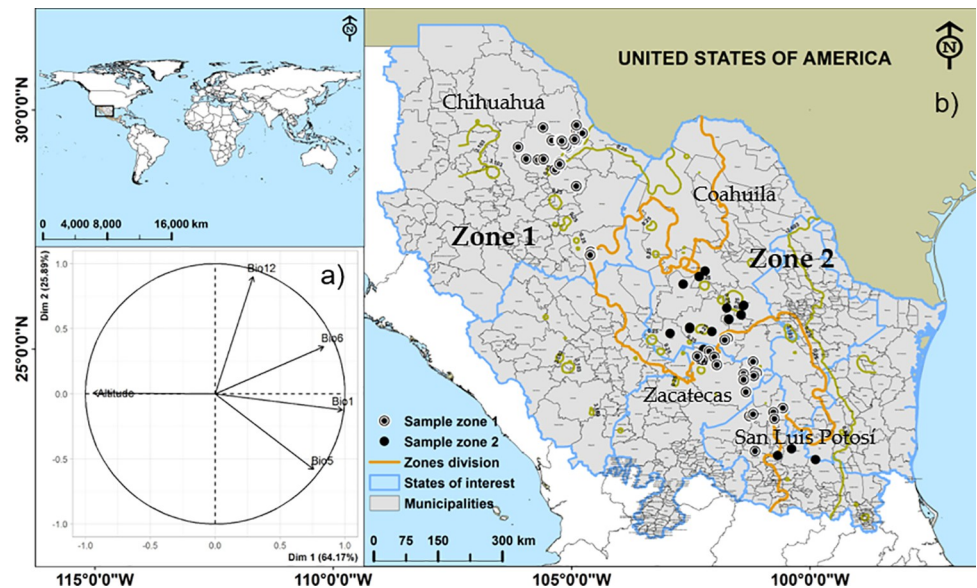


Fig 1. Location of the study area in the global context (a), two bioclimatic zones generated of *Agave lechuguilla* Torr. using principal components analysis (PCA) without rotation, where brown lines indicate the limits of the bioclimatic zones; circles indicate the sampling sites of the species; while the polygons (blue lines) indicate the states of the country, and thin black lines indicate municipal boundaries (b). Correlation plot of bioclimatic variables and altitude with eigenvectors of components 1 and 2 of the principal components analysis (c). [Shapefiles were taken from: world map [34], State Political Division of Mexico [35] and Municipal Political Division of Mexico [36]. Fig 1A may be similar, but not identical to any published figure, however, in this paper it is used solely for illustrative purposes, no information or databases were extracted from it.

<https://doi.org/10.1371/journal.pone.0274641.g001>

than in Zone 1. This appears to obey basically minimum temperature (Bio 6) and, to a lesser degree, precipitation (Bio 12), which are also the highest (Fig 2, S2 Table) in this zone. There was no proof for rejecting the H_0 that mean annual precipitation and altitude are the same among zones ($p = 0.956$ and $p = 0.106$).

3.2. Characterization of the *Agave lechuguilla* Torr. sample

The total number of plants evaluated was 499, distributed geographically over a distance of more than 1000 km, according to Fig 1: (Chih = 146, Coah = 119, SLP = 124 and Zac = 110).

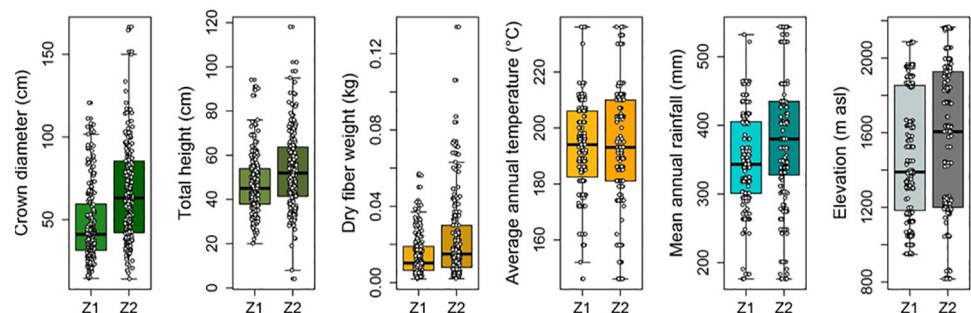


Fig 2. Boxplot with Jitter of dendrometric (first three plots) variables of *Agave lechuguilla* Torr. and values of bioclimatic variables obtained from each sampling site of individuals of this species (last three plots). Where Z1 = Zone 1 and Z2 = Zone 2, represents the bioclimatic zones generated through principal components analysis (Fig 1).

<https://doi.org/10.1371/journal.pone.0274641.g002>

Table 1. Descriptive statistics of the sample used to predict dry fiber in *Agave lechuguilla* Torr.

Zone	Var	Mean	MeanCI'	0.05	0.10	0.25	Median	0.75	0.90	0.95	Range	SD	VCoef	MAD	IQR
Zone 1 (n = 200)	Cd	46.82	±3.20	19.98	23.00	31.50	41.00	59.25	83.60	94.53	106.00	22.93	0.49	17.42	27.75
	H	47.03	±1.88	28.95	32.00	38.00	45.00	54.00	62.83	71.10	74.00	13.49	0.29	11.86	16.00
	Dfw	14.19	±1.49	3.90	5.00	6.45	10.20	19.00	29.73	34.15	55.20	10.66	0.75	7.12	12.55
Zone 2 (n = 199)	Cd	66.10	±4.05	28.45	32.90	42.00	63.00	85.43	102.10	114.70	152.50	28.99	0.44	31.88	43.43
	H	52.76	±2.42	29.09	33.00	41.40	52.00	63.70	74.04	83.00	114.00	17.28	0.33	16.31	22.30
	Dfw	22.21	±2.87	4.00	5.00	8.00	15.00	30.00	0.05	46.6	132.00	20.48	0.92	13.34	22

Cd = crown diameter (cm), H = total height (cm), Dfw = dry fiber weight (g), MeanCI' = confidence interval for the mean, 0.05, . . . , 0.95 = quantiles of the variable, SD = standard deviation, VCoef = coefficient of variation, MAD = median absolute deviation, IQR = interquartile range.

<https://doi.org/10.1371/journal.pone.0274641.t001>

The plants were obtained from 20 municipalities and 46 ejidos (S1 Appendix 1). It can be seen that *A. lechuguilla* individuals are taller (52.76 cm) and crown diameters are larger (66.10 cm) in Zone 2 (Table 1, Fig 1), and therefore, they accumulate on average a larger amount of fiber in their plant tissues (22.21 g plant⁻¹), but up to maximums of 34.15 and 46.60 g plant⁻¹, according to percentile 95 (Zone 1 and Zone 2). In general, to predict *A. lechuguilla* dry fiber, a selective sampling is applied, considering a lower limit for plant height; for example, in Zone 2 of this study, Blando and Baca [12] selected plants ≥ 45 cm tall finding that the accumulation of dry fiber can be between 45.6 and 70.5 g plant⁻¹, much higher than the average of our study. We included plants ≥ 10 cm tall, and for this reason, when averages (or even medians) are compared with previous studies, the differences are contrasting.

3.3. Analysis of predictors of *Agave lechuguilla* Torr. dry fiber

In a preliminary analysis Eq 2 was adjusted with each predictor (Cd, H, Cd×H). Although all show the capacity to predict *A. lechuguilla* dry fiber ($0.519 < R^2 < 0.633$ and $p\text{-value} < 2e-16$), only Cd passes all the assumptions of a linear model (Table 2). For this reason, Cd was selected as the predictor of *A. lechuguilla* dry fiber. It has been demonstrated that the best predictors in this species are Cd and H, e.g. Valencia *et al.* [37] used Cd (cm) and plant volume (m³) to predict plant fresh weight (kg), with correlations of 0.9272 and 0.9433, respectively. Flores *et al.* [26] show that Cd and H efficiently predict aboveground biomass, explaining 91.4%. With destructive methods, it has been proven that heart volume (cm³) can explain from 57.9% [7, 25] to 90.98% [11] of the heart fresh weight. The moisture index (MPa) and the index of photo-synthetically active radiation (mmol m²·S⁻¹) explain 97% and 25% of the biomass of this species [38].

In addition, statistically significant ($p < 0.001$) results were obtained with the hypothesis tests on the coefficients of regression with different sample sizes (Table 3): intercept (β_0), slope [$\ln Cd$ (β_1)], dummy variable [Zone (β_2)] and interaction [$\ln Cd$:Zone (β_3)] of Eq 3. This demonstrates correct bioclimatic zoning of *A. lechuguilla* productivity with the PCA and shows that the predictor ($\ln Cd$) of dry fiber does not originate from a random model.

3.4. Model for predicting *Agave lechuguilla* Torr. dry fiber

The final model obtained with ML showed an R^2 of 0.629. The coefficients of regression were highly significant, $p < 0.05$ (Table 4), similar to the average obtained with different sample sizes (Table 3). The dummy variable (Zone) and the interaction $\ln Cd$:Zone show p values of 0.0004 and 0.0014 (Table 4), giving way to an equation for each bioclimatic zone (Fig 3A) for estimating *A. lechuguilla* dry fiber.

Table 2. Statistics of fit and verification of assumptions of the power model in logarithmic form of predictors of dry *Agave lechuguilla* Torr. fiber weight.

Coeff.	Estimate	S.E.	t-value	Pr(> t)	R ²	KS	BP	D-W
Intercept (β_0)	-9.132	0.186	-49.20	< 2e-16 ***	0.577	0.233	0.223	0.088
$\ln Cd$ (β_1)	1.226	0.047	26.03	< 2e-16 ***				
Intercept (β_0)	-11.027	0.289	-38.09	< 2e-16 ***	0.519	2.2e-16	3.58e-16	0.004
$\ln H$ (β_1)	1.733	0.750	23.18	< 2e-16 ***				
Intercept (β_0)	-10.721	0.219	-48.92	< 2e-16 ***	0.633	0.003	0.202	0.102
$\ln CdH$ (β_1)	0.821	0.028	29.26	< 2e-16 ***				

Coeff = coefficient, S.E. = standard error, R² = coefficient of determination. Value of p of the normality test, KS (Kolmogorov-Smirnov), Variance homogeneity, BP (Breusch Pagan) and independence, D-W (Durbin Watson). *Cd* = crown diameter, *H* = total plant height, *ln* = natural logarithm, β_0 and β_1 = coefficients of regression. Signif. codes: 0

*** 0.001

** 0.01

* 0.05 ' ' 0.1 ' ' 1.

<https://doi.org/10.1371/journal.pone.0274641.t002>

Because of the wide variation in fiber observed in this species (Table 1), in the regression analysis we present some influential observations (Fig 3B) on the model coefficients: D_i (DFFITS), but none with $D_i > 1$ that could be eliminated. There were no influential observations regarding DFBETAS. With COVRATIO, several observations were outstanding as influential, but we underline that no atypical r_i data were detected. The correction (1.127) and multiplicative factor obtained at the final equation can correct the bias and improve the estimations of *Dfw* by 12.7%.

In this way, the resulting equation for estimating *A. lechuguilla* fiber dry weight in bioclimatic zones 1 and 2, respectively (Fig 1B), are the following:

$$Dfw = \exp(-8.82605 + 1.16201 \times (\ln Cd)) \times F.C. \quad (4)$$

$$Dfw = \exp(-8.82605 + (-1.4682) + (0.33514 + 1.16201) \times \ln Cd) \times F.C. \quad (5)$$

where: *Dfw* = fiber dry weight (kg), *lnCd* = natural logarithm of crown diameter (cm), exp = exponential, *F.C.* = correction factor (1.127).

In accord with the confidence level established (95%), the studentized residuals of the model for predicting *A. lechuguilla* dry fiber present homoscedasticity ($p = 0.1295$) and are normal ($p = 0.5343$) and independent ($p = 0.254$) (Fig 4A–4C, respectively), assuring a robust, unbiased and efficient model [15, 28, 29].

Table 3. *p* values for the hypothesis tests of intercept, slope and interactions with the dummy variable at different sample sizes, with the power model in logarithmic form to predict *Agave lechuguilla* Torr. dry fiber.

Coefficient	Sample size (%)						Mean (coef.)	S.D. (coef.)
	100	90	80	70	60	50		
Intercept (β_0)	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	-8.745	0.119
$\ln Cd$ (β_1)	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	1.139	0.031
Zone (β_2)	1.0e-4	6.2e-5	0.001	1.2e-4	6.6e-5	5.4e-5	-1.375	0.236
$\ln Cd:Zone$ (β_3)	0.020	0.003	0.015	0.002	0.002	9.2e-4	0.314	0.062
R	0.771	0.764	0.781	0.778	0.757	0.782		

Cd = crown diameter, *ln* = natural logarithm, $\beta_0, \dots, \beta_3$ = coefficients of regression, *Zone* = dummy variable, Mean (coef.) = mean of regression coefficients, S.D. (coef.) = standard deviation of regression coefficients. R = Coefficient of multiple correlation.

<https://doi.org/10.1371/journal.pone.0274641.t003>

Table 4. Fit statistics of the power-dummy model in logarithmic form for estimating *Agave lechuguilla* Torr. fiber dry weight in arid and semiarid regions of Mexico.

Coefficients	Estimate	Std. error	t-value	Pr(> t)	I.C. 2.5%	I.C. 97.5%
Intercept (β_0)	-8.826	0.252	-35.086	< 2e-16 ***	- 9.321	-8.332
$\ln Cd$ (β_1)	1.162	0.066	17.640	< 2e-16 ***	1.033	1.292
Zone (β_2)	-1.468	0.414	-3.546	0.0004 ***	- 2.282	-0.654
$\ln Cd:Zone$ (β_3)	0.335	0.104	3.219	0.0014 **	0.130	0.540

I.C. = confidence interval, Cd = crown diameter, $\ln$ = natural logarithm, $\beta_0, \dots, \beta_3$ = coefficients of regression, $Zone$ = dummy variable, Signif. codes: 0

*** 0.001 ** 0.01 * 0.05 . 0.1 . 1.

<https://doi.org/10.1371/journal.pone.0274641.t004>

3.5. Model capacity for predicting dry fiber content of *Agave lechuguilla* Torr.

Model validation using machine learning, with 100 independent data (data that the model would never have “seen”) showed that the model exhibits good capacity for predicting *A. lechuguilla* fiber dry weight. The fit statistics (all in logarithmic units) in the four methods of validation (R^2) are clearly lower (e. g. average R^2 of 0.544) than those resulting from the set of training data ($R^2 = 0.629$), but very similar to each other (Table 5). This guarantees that, by using the generated equations to predict *A. lechuguilla* dry fiber with data that the model has never seen, the estimations are adequate.

According to the literature, using independent data to validate a model is one of the most recommended techniques [39, 40]. Using a set of independent data and applying the model generated with the training set, Velasco *et al.* [3] evaluated its capacity to predict dry fiber in this species and found that the fit statistics and predictions are very similar, assuring robustness when predicting with new data.

In similar studies conducted in Zone 2 (Fig 1) with $n = 287$, Blando and Baca [12] found that Cd and number of usable leaves explained 76.09% of the *A. lechuguilla* fiber. Using 95 individuals ($H \geq 25$ cm) in Zone 2 of this study, Pando *et al.* [13] demonstrated that H and Cd (as a single independent variable) explained 86.9% of *A. lechuguilla* fiber dry weight, with an error of 5.041 g. In the same Zone ($n = 240$) Velasco *et al.* [3] found that the diameter and height of the bud explained 68% of the fiber in *A. lechuguilla*. It is evident that the fit statistics of each model differ from that of our study and from each other due to the sample size (n), sampling method (directed, selective, quadrants), plant selection (minimum height),

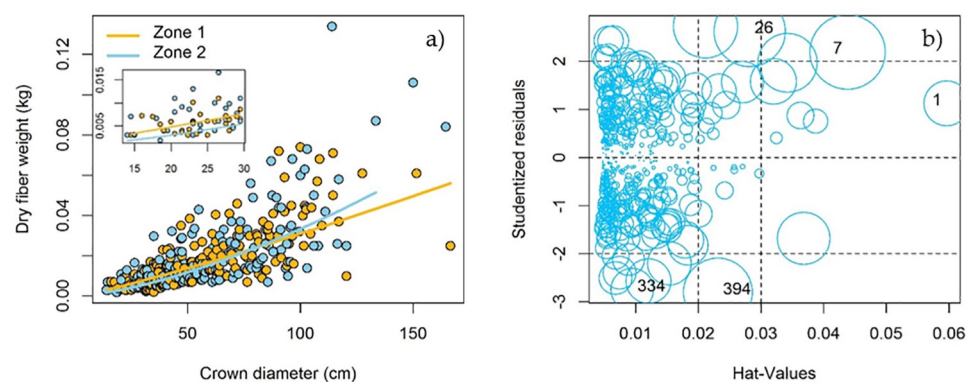


Fig 3. Observed data of dry fiber of *Agave lechuguilla* Torr. in bioclimatic zone 1 (brown circles) and bioclimatic zone 2 (blue circles) and data estimated with the dummy model in each bioclimatic zone (brown and blue lines, respectively) (a), analysis and representation of influential observations of the model (b).

<https://doi.org/10.1371/journal.pone.0274641.g003>

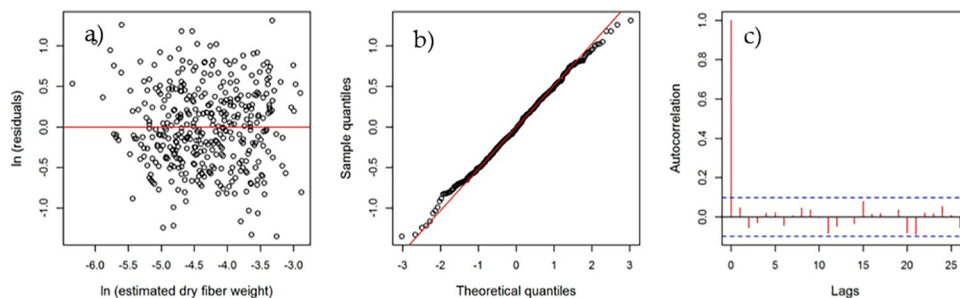


Fig 4. Analysis of residuals of the dummy model for estimating *Agave lechuguilla* Torr. fiber dry weight in northeastern Mexico. Variance homogeneity (a), normality (b), and independence (c).

<https://doi.org/10.1371/journal.pone.0274641.g004>

geographic location of the study (zone 1 or 2), model type (linear, non-linear, simple or multiple), type of predictor and even compliance or non-compliance with the model assumptions.

Although linear models have been generated to estimate fiber weight in *A. lechuguilla* [12, 13], the allometry reveals exponential-type changes in dimensions relative to the parts, e. g. in *Agave salmiana* Otto ex. Salm. ssp. crassispina (Trel.) Gentry aboveground biomass follows this pattern [41], as in *Agave salmiana* ssp. Crassispina [42], and as has been demonstrated for prediction of aboveground biomass in *A. lechuguilla* [26] and dry fiber in our study, applying the power equation (Eq 1) in its linear form (Eq 2).

3.6. Canonical relationship between *A. lechuguilla* dry fiber and bioclimatic variables

According to the first and second canonical variable, the degree of correlation between dendrometric variables (matrix *Y*) is higher ($r = 0.9000$) than that between bioclimatic variables (matrix *X*), which is $r = 0.7345$ (Fig 5A). The correlation in the first three dimensions is 0.8101, 0.5395 and 0.2510, respectively. Crossed canonical correlation shows that the minimum temperature of the coldest month, Bio 6, and annual precipitation, Bio 12 (Fig 5A, 5B and Table 6), correlated positively with *A. lechuguilla* dry fiber production in these bioclimatic zones. According to the three models of general circulation, CCSM4, HadGEM2-AO and MIROC5 [20], in a scenario of moderate CO₂ emissions (4.5 W/m²), it is expected that by 2050, Bio 6 will decrease 0.56°C in the area of *A. lechuguilla* distribution, possibly affecting the species' fiber production, while Bio 12, will decrease only 1.79%.

4. Conclusions

The principal components analysis generated two bioclimatic zones, in Zone 2 (located in the eastern part of the study area) *A. lechuguilla* productivity is higher than in Zone 1. This zoning

Table 5. Fit statistics of the evaluation of the predictive capacity of the model for predicting *Agave lechuguilla* Torr. fiber dry weight, using machine learning with an independent sample ($n = 100$).

Validation	Set data	n	R ²	RMSE	MAE
Model fit	Training	399	0.629	0.487	0.397
Leave One Out Cross-Validation (LOOCV)	Test	100	0.512	0.574	0.461
k-fold validation (Cross-Validation)	Test	100	0.534	0.564	0.462
Repeated K-fold (Cross-Validation)	Test	100	0.545	0.559	0.459
Bootstrap Validation	Test	100	0.583	0.561	0.453

n = sample size, R² = coefficient of determination, RMSE = root mean square error, MAE = mean absolute error.

<https://doi.org/10.1371/journal.pone.0274641.t005>

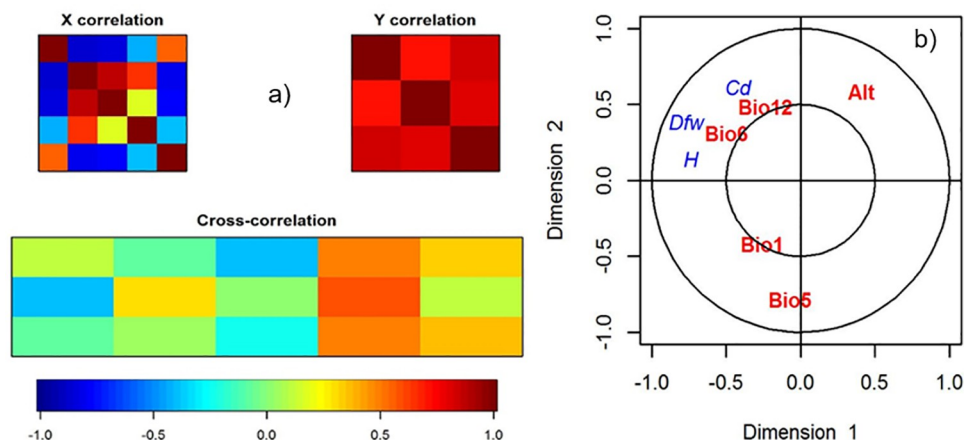


Fig 5. Canonical correlation analysis (CCA) between bioclimatic variables [matrix X: Bio 1 (°C), maximum temperature of the hottest month, Bio 5 (°C), minimum temperature of the coldest month, Bio 6 (°C), and mean annual precipitation, Bio 12 (mm) and altitude (m)]; and *Agave lechuguilla* Torr. dendrometric variables [matrix Y: crown diameter (Cd, cm), total plant height (H, cm), fiber dry weight (Dfw, kg)] (a) and their expression in the first two dimensions of the CCA.

<https://doi.org/10.1371/journal.pone.0274641.g005>

Table 6. Correlation analysis between groups of dendrometric variables of *Agave lechuguilla* Torr. and bioclimatic variables.

	Cd	H	Dfw	Alt	Bio 1	Bio 5	Bio 6	Bio 12
Cd	1	0.0006	0.00	0.6703	0.7888	0.0758	0.0292	0.1592
H	0.70	1	0.00	0.0940	0.2068	0.9075	0.0075	0.6685
Dfw	0.84	0.80	1	0.7177	0.8874	0.2884	0.0258	0.1070
Alt	0.10	-0.38	-0.09	1	0.0000	0.0000	0.0557	0.0128
Bio 1	-0.06	0.29	0.03	-0.86	1	0.0000	0.0017	0.0000
Bio 5	-0.41	0.03	-0.25	-0.84	0.86	1	0.4456	0.0000
Bio 6	0.49	0.58	0.50	-0.43	0.66	0.18	1	0.0802
Bio 12	0.33	0.10	0.37	0.55	-0.81	-0.76	-0.40	1

Cd = crown diameter (cm), H = total height (cm), Alt = altitude (m), Dfw = fiber dry weight (kg), Bio 1 = mean annual temperature (°C), Bio 5 = maximum temperature of the warmest month (°C), Bio 6 = minimum temperature of the coldest month (°C), Bio 12 = annual rainfall (mm).

<https://doi.org/10.1371/journal.pone.0274641.t006>

is useful when focusing efforts specifically on management and use of this species. The best predictor of dry fiber content is crown diameter (Cd). Using dummy regression, it was possible to establish an equation for each bioclimatic zone (Fig 1B). Validation of this equation revealed high predictive capacity of the model, which is also easy, rapid, and inexpensive to use and is useful in large part of the known distribution area of the species. The results suggest that the species could be vulnerable to climate change with a decrease in Bio 6 of more than 0.5°C by 2050 and there would likely be a gradual reduction in *A. lechuguilla* fiber production over the coming years.

Supporting information

S1 Table. *Agave lechuguilla* Torr. individuals evaluated and ejidos sampled by municipality and state.

(DOCX)

S2 Table. Fit statistics of the power-dummy model in logarithmic form to establish differences in *Agave lechuguilla* Torr. dendrometric variables and bioclimatic variables between zones derived from the PCA.

(DOCX)

Acknowledgments

To the National Council of Science and Technology (CONACYT) for the scholarship awarded to the first author for postgraduate studies.

Ethics statement: All necessary permits were obtained for the described study, which complied with all relevant regulations. This protocol was presented to the National Council for Science and Technology (CONACyT), represented by Rafael Pando Cerón, technical secretary, and to the National Forestry Commission (CONAFOR) represented by Tania Ananí Limón Magaña Administrative Secretary, for review and if applicable approval; project that was finally authorized and financed through project number 2017-4-29267. Field access to the study areas was authorized by the landowners. None of the species sampled in this study are considered protected species, hence they are being exploited through management programs.

Author Contributions

Conceptualization: José Óscar M. López-Díaz, Jorge Méndez-González, Emily García-Montiel, Alejandro Zermeño-González, José J. Corral Rivas.

Data curation: Fátima M. Méndez-Encina, Rocío Mendieta-Oviedo, Andrés Flores, Emily García-Montiel, Alejandro Zermeño-González.

Formal analysis: Félix de J. Sánchez-Pérez.

Funding acquisition: Pablito M. López-Serrano, José J. Corral Rivas.

Investigation: José Óscar M. López-Díaz, Félix de J. Sánchez-Pérez, Librado Sosa-Díaz, Alejandro Zermeño-González.

Methodology: José Óscar M. López-Díaz, Jorge Méndez-González, Pablito M. López-Serrano, Félix de J. Sánchez-Pérez, Librado Sosa-Díaz, Andrés Flores, Víctor H. Cambrón-Sandoval.

Supervision: Fátima M. Méndez-Encina, Rocío Mendieta-Oviedo, Librado Sosa-Díaz, Emily García-Montiel, Alejandro Zermeño-González.

Validation: Pablito M. López-Serrano, Félix de J. Sánchez-Pérez, Rocío Mendieta-Oviedo, Víctor H. Cambrón-Sandoval.

Visualization: Fátima M. Méndez-Encina, Andrés Flores, Víctor H. Cambrón-Sandoval.

Writing – original draft: José Óscar M. López-Díaz.

Writing – review & editing: José Óscar M. López-Díaz, Jorge Méndez-González, Fátima M. Méndez-Encina, Emily García-Montiel.

References

1. Briones O, Búrquez A, Martínez-Yrizar A, Pavón N, Perroni Y. Biomasa y productividad en las zonas áridas mexicanas. *Madera y Bosques*. 2018; 24: 1–19. <https://doi.org/10.21829/myb.2018.2401898>
2. Rzedowski J. Vegetación de México. Ciudad de México: Limusa; 1978.
3. Velasco BE, Arredondo GA, Zamora-Martínez MC, Moreno SF. Modelos predictivos para la producción de productos forestales no maderables: Lechuguilla. Manual Técnico Num. 2: INIFAP; 2009.

4. Castillo QD, Cano PA, Berlanga RCA. Establecimiento y aprovechamiento de lechuguilla (*Agave lechuguilla* Torr.). 1st ed. México: Comisión Nacional Forestal-Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias; 2012. <https://doi.org/10.24850/j-tyca-2017-03-04>
5. Marroquín JS, Borja LG, Velásquez CR, De la Cruz CJA. Estudio Ecológico dasonómico de las zonas áridas del norte de México. 2a ed. Ciudad de México: INIF; 1981.
6. SEMARNAT. Norma Oficial Mexicana NOM-008-SEMARNAT-1996. SEMARNAT: México; 1996.
7. Berlanga RCA, González LLA, Franco LH. Metodología para la evaluación y manejo de lechuguilla en condiciones naturales. Folleto Técnico Núm. 1. Saltillo, Coahuila, México: Campo Experimental "La Saucedá" CIRNE-INIFAP; 1992.
8. Castillo QD, Mares AO, Villavicencio GEE. Lechuguilla (*Agave lechuguilla* Torr.), planta suculenta de importancia económica y social de las zonas áridas y semiáridas de México. Boletín de la Sociedad Latinoamericana y del Caribe de Cactáceas y otras Suculentas. 2011; 8: 6–9.
9. Mayorga-Hernández E, Rössel-Kipping D, Ortiz-Laurel H, Quero-Carillo AR, Amante-Orozco A. Análisis comparativo en la calidad de fibra de *Agave lechuguilla* Torr., procesada manual y mecánicamente. Agrociencia. 2004; 38: 219–225.
10. Castillo QD, Berlanga RCA, Pando MM, Cano PA. Regeneración del cogollo de *Agave lechuguilla* Torr. de cinco procedencias bajo cultivo. Revista Ciencia Forestal en México. 2008; 33: 27–40.
11. Pando-Moreno M, Eufrazio O, Jurado E, Estrada E. Post-Harvest growth of lechuguilla (*Agave lechuguilla* Torr., Agavaceae) in northeastern Mexico. Economic Botany. 2004; 58: 78–82. [https://doi.org/10.1663/0013-0001\(2004\)058\[0078:pgola\]2.0.co;2](https://doi.org/10.1663/0013-0001(2004)058[0078:pgola]2.0.co;2)
12. Blando-Navarrete JL, Baca MS. Determinación del potencial productivo de la lechuguilla (*Agave lechuguilla* Torr) en el municipio de San Juan de Guadalupe, Dgo. Revista Chapingo Serie de Zonas Áridas. 2001; 2: 100–105. <https://doi.org/10.29298/rmcf.v3i9.531>
13. Pando-Moreno M, Pulido R, Castillo D, Jurado E, Jiménez J. Estimating fiber for lechuguilla (*Agave lechuguilla* Torr., Agavaceae), a traditional non-timber forest product in Mexico. Forest Ecology and Management. 2008; 255: 3686–3690. <https://doi.org/10.1016/j.foreco.2008.02.053>
14. Blando NJL, Baca MS, Martínez SM. Caracterización ecológica y socioeconómica de la lechuguilla (*Agave lechuguilla* Torrey). In: Martínez S.M., Ecología y usos de especies forestales de interés comercial de las zonas áridas de México. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. 2013. pp. 70–90.
15. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. 4th ed. Hoboken, NJ, USA: John Wiley & Sons; 2012.
16. Hourri A, Machaka-Houri N. *Agave lechuguilla* as a potential biomass source in arid areas. Journal of Forestry Development of Energy, Water and Environment Systems. 2016; 4: 89–93. <https://doi.org/10.13044/j.sdwes.2016.04.0008>
17. Castillo QD, Berlanga RCA, Cano PA. Establecimiento y manejo de una plantación de lechuguilla. Campo Experimental CIR-NORESTE: INIFAP; 2005.
18. Nobel PS, Quero E. Environmental productivity indices for a Chihuahuan Desert CAM plant, *Agave lechuguilla*. Ecology. 1986; 67: 1–11. <https://doi.org/10.2307/1938497>
19. Flores DMP. Bordeando el monte: la lechuguilla. Un recurso olvidado. Secretaría del Medio Ambiente; 2018.
20. Karger DN, Conrad O, Böhrer J, Kawohl T, Kreft H, Soria AR, et al. Climatologies at high resolution for the earth's land surface areas. Sci. Data. 2017; 4: 1–20. Available from: <https://doi.org/10.1038/sdata.2017.122> PMID: 28872642
21. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. 2021. Available from: <https://www.r-project.org/>
22. Vaissie P, Monge A, Husson F. FactoShiny: Perform Factorial Analysis from 'FactoMineR' with a Shiny. R Package Version 2.4. 2021. 18. Available from: <https://cran.r-project.org/web/packages/Factoshiny/Factoshiny.pdf>
23. Husson F, Josse J, Le S, Mazet J. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R Package Version 2.4. 2020. 100. Available from: <https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>
24. Wood AG. A potential bias in log-transformed allometric equations. Wader Study Group Bull. 1986; 47: 17–19.
25. Zárate LA, Franco H, Berlanga CA. Análisis dimensional en lechuguilla. In: Proceedings of the III Simposio Nacional sobre Ecología, Manejo y Domesticación de Plantas Útiles del Desierto. INIFAP; 1991. Pp. 13–14.

26. Flores HC de J, Méndez GJ, Sánchez PF de J, Méndez EFM, López DÓM, López S PM. Allometric equations for predicting *Agave lechuguilla* Torr. aboveground biomass in México. *Forest*. 2020; 784: 1–17. <https://doi.org/10.3390/f11070784>
27. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. Caret: Classification and Regression Training. R Package Version 6.0–88, 2021. 224. Available from: <https://cran.r-project.org/web/packages/caret/caret.pdf>
28. Faraway JJ. *Linear Models with R*. 2nd ed. Boca Raton, FL, USA: CRC press; 2014.
29. Fox J. *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Thousand Oaks, CA, USA: Sage Publications; 2016.
30. Sprugel DG. Correcting for bias in log-transformed allometric equations. *Ecology*. 1983; 64: 209–210. <https://doi.org/10.2307/1937343>
31. Déjean S, González I. CCA: Canonical Correlation Analysis. R Package Version 1.2.1. 2021. Available from: <https://cran.r-project.org/web/packages/CCA/CCA.pdf>
32. Castillo-Quiroz D, Martínez-Burciaga OU, Ávila-Flores DY, Castillo-Reyes F, Sánchez-Chaparro JD. Identification of potential areas for establishment of plantations of *Agave lechuguilla* Torr. in Coahuila, Mexico. *Open Journal of Forestry* 2014; 4: 520–526. <https://doi.org/10.4236/ojfor.2014.45056>
33. Martínez SM, Hermosillo RDE, Mojica GAS, Prieto AJA. Potencial productivo y zonificación para el uso y manejo de especies forestales de zonas áridas. *Campo Experimental La Campana*, Chihuahua, México: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias; 2015.
34. Efrain Maps. Shapefile (*.shp) del mundo. Carlos Efraín Porto Tapiquén Geografía, SIG y Cartografía Digital. Valencia, España. 2020. <https://www.efrainmaps.es/descargas-gratuitas/mundo/>.
35. CONABIO. División política estatal 1:250000. 2015. Catálogo de metadatos geográficos. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. 2016: metadato 19-04-2016. http://www.conabio.gob.mx/informacion/metadatos/gis/dest_2015gw.xml?_httpcache=yes&_xsl=/db/metadatos/xsl/fgdc_html.xsl&_indent=no.
36. CONABIO. División política municipal 1:250000. 2018. Catálogo de metadatos geográficos. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. 2018: metadato 04-12-2018. http://www.conabio.gob.mx/informacion/metadatos/gis/muni_2018gw.xml?_httpcache%20=%20yes&_xsl=/db/metadatos/xsl/fgdc_html.xsl&_indent%20=%20no.
37. Valencia CCM, Sánchez MH, Quiñones VJJ, Martínez RJJ. Predicción de la biomasa aérea de cuatro especies arbustivas del semidesierto en la comarca lagunera. *Agrofaz*. 2003; 3: 365–372.
38. Quero E, Nobel PS. Predictions of field productivity for *Agave lechuguilla*. *Journal of Applied Ecology*. 1987; 24: 1053–1062. <https://doi.org/10.2307/2404001>
39. Snee RD. Validation of regression models: methods and examples. *Technometrics*. 1977; 19: 415–428.
40. Arboretti G R, Salmaso L. Model performance analysis and model validation in logistic regression. *Statistica*. 2003; 43: 375–396.
41. García-Moya E, Romero-Manzanares A, Nobel PS. Highlights for Agave productivity. *Global Change Biology Bioenergy*. 2011; 3: 4–14.
42. Martínez-Morales R, Meyer SE. A demographic study of maguey verde (*Agave salmiana* ssp. *Crassispina*) under conditions of intensive utilization. *Desert Plants*. 1985; 7: 61–66.