

New approach for modeling climatic data with applications in modeling tree species distributions in the states of Jalisco and Colima, Mexico

R.M. Reich^{a,*}, C. Aguirre-Bravo^b, V.A. Bravo^a

^aDepartment of Forest, Rangeland and Watershed Stewardship, Colorado State University, Fort Collins, CO 80525, USA

^bRocky Mountain Research Station, USDA Forest Service at the Rocky Mountain Research Station, 240 West Prospect Rd., Fort Collins, CO 80526, USA

Received 8 January 2007; received in revised form 26 November 2007; accepted 14 February 2008
Available online 3 April 2008

Abstract

Spatial models of monthly climatic data in the states of Jalisco and Colima, Mexico, are developed using a combination of satellite imagery, topographic data and climatic data from 256 weather stations. The models accounted for 45–85% of the variability in the monthly temperature, precipitation and evaporation. In spite of having highly skewed distributions, cross-validation showed the models to have nominal prediction bias. The monthly climatic models for temperature, precipitation and evaporation were used to define 12 climate zones. Comparing the climatic zones against observed patterns of vegetation showed that the model captured the general placement of arid, semi-arid, temperate and tropical dry forests. Distribution models for four tree species are derived based on climatic constraints controlling their abundance. Site conditions are spatially simulated using Geographic Information Systems and represent resource gradients (temperature, precipitation, evaporation) and indirect variables, which reflect soil properties (topographic data). The application of the climatic zones in modeling the spatial distribution of vegetation types, soil texture and other related properties, identifying habitats suitable for selected endangered and threatened species and land use planning are discussed.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Climate zones; Cross-validation; Error estimation variance; Species distribution models; Topical dry forests

1. Introduction

The forests of Jalisco and Colima, Mexico contain a very diverse and unique community of endemic and specialized species of plants, animals, reptiles and amphibians. The tropical dry forests, which are located along the Pacific coast are among the richest tropical dry forests in the world, and have more endemic species than dry forests elsewhere in the neotropics (Challenger, 1998). At higher elevations the forests slowly change to pine–oak forests which are host to a number of endemic species and the region is recognized as a center of diversification for the genera *Quercus* (Nixon, 1993).

*Corresponding author. Tel.: +1 970 491 6980.

E-mail addresses: robin@warnercnr.colostate.edu (R.M. Reich), caguirrebravo@fs.fed.us (C. Aguirre-Bravo), vbravo@lamar.colostate.edu (V.A. Bravo).

These forests have been extensively exploited through logging and agricultural activities. Habitat destruction not only affects the vegetation and wildlife, but also influences the local climate. Changes in vegetation cover can affect local precipitation and temperature patterns. Vegetation patterns and soil composition can influence cloud formation and precipitation through their impact on evaporation and convection (de Sherbinin, 2002).

The climate of a region is determined by long-term variability in temperature, precipitation and evaporation. In semi-arid regions such as Jalisco and Colima, precipitation is an important resource. Any fluctuations in climatic patterns could influence precipitations patterns as well as other natural processes (e.g. droughts, fires, hurricanes, floods, insects and diseases outbreaks) that are of vital significance to social and economic systems (Allen and Hoekstra, 1992) as well as ecological systems (Bailey, 1996).

For many regional and local applications, users of climate models have expressed concern with the differences in the scale of the output from global climate models (hundreds of km) (Cohen, 1990; Gates, 1985; Lamb, 1987) and the scale required for assessing climatic impacts (IPCC, 1994). The availability of spatially explicit fine-scale climatic models is essential for development of policies and subsequent resource management strategies (Bailey, 2002). In turn, assessments can be made of the capability and suitability of the land for various kinds of use and resource management applications. Scientists then may use this information to study and model fine-scale responses of ecological systems to various inputs of change.

There are a number of considerations that must be taken into account when modeling climatic data at fine spatial resolutions. Changes in topography and vegetation, for example, may significantly affect local climatic responses (Bonan, 2002; Schultz, 1995), and their spatial–temporal variability (fine-scale). These changes in fine-scale spatial dependencies are central to the development of precise geospatially explicit climatic syntheses. When data are sparse over a region or have weak spatial dependency, the underlying assumptions about variation among data points may differ and the choice of interpolation method and parameter estimation may become critical. No single method of spatial interpolation is optimal for all situations, and thus it is important to compare results obtained from various methods (Nalder and Wein, 1998).

The analysis and interpretation of spatial data is an important part of geostatistics, and as with any type of statistical analysis is highly dependent on the individual performing the analysis. In general, judgment and experience play an important role in selecting the proper spatial interpolation technique for each individual case (Englund, 1990). This is partly due to the variety of available spatial interpolation methods which range from simple intuitive predictions to more sophisticated and complex procedures (Cressie, 1991).

There are techniques available for generating fine scale climatic information, but some tend to be complex and/or computationally expensive (IPCC, 1994). Although there have been numerous articles written on spatial interpolation, there is little or no agreement among authors on the superiority of one technique over others. Additionally, the increased interest in Geographic Information Systems (GIS) has resulted in many of these interpolation techniques being embedded in GIS systems to create a set of ready-to-use spatial interpolation techniques (Burrough, 1986; Heine, 1986; Oliver, 1990; Royle et al., 1981).

In a recent study Reich et al. (2004) described a relatively quick and economical method of accurately modeling the spatial distribution of forest stand structure using field data, topographic data and satellite imagery. These procedures are used to develop prediction maps of average monthly temperature, precipitation and evaporation at a 30 m spatial resolution to support ongoing efforts to monitor and assess conditions and change of land resources in the states of Jalisco and Colima, Mexico (Schreuder et al., 2003). Probability indices were developed from the prediction maps to define 12 climatic zones that could be useful in studying impacts of climatic change on the species distributions. To demonstrate the application of the climatic models, the distribution of four selected tree species are modeled by analyzing environmental characteristics (topographic data and climatic data) at their known locations.

2. Methods

2.1. Study area

This study was conducted in the states of Jalisco and Colima, which are located in west central Mexico and covers an area of approximately nine million hectares (20 million acres). Although Jalisco is larger in area

(93%), the state of Colima (7%) has a distinctive role in the economy of the region. Four major ecological regions provide natural resources and environmental conditions that make this region one of the most prosperous in Mexico. Ecoregions include the transversal neo-volcanic system, the southern Sierra Madre, the Southern and Western Pacific Coastal Plain and Hills and Canyons, and the Mexican High Plateau. Linked to these ecological regions, are several important hydrological regions (watersheds) that drain to the Pacific Ocean (Lerma-Santiago, Huicicila, Ameca, Costa de Jalisco, Armeria-Coahuayana, Balsas, and El Salado). One of the watersheds, the Lerma-Santiago Hydrological Region, is connected to Chapala Lake, the most important source of water for the City of Guadalajara (INEGI, 2001).

The region's biophysical heterogeneity is reflected in a large diversity of plant and animal species. Within this region there are a significant number of species of mammals and birds, many of which are threatened by human activities. Some of the plant and animal species are endemic to specific locations. Examples are the areas of pine–oak forest that are home to “specialty” birds such as the thick-billed parrot, the Mexican-spotted owl and woodpeckers.

2.2. Climate data

Climatic data from 256 weather stations were obtained from Mexico's INIFAP (National Institute for Forestry and Agriculture Research) for use in this study (Fig. 1). INIFAP manages and maintains one of the largest climatic databases in Mexico for agriculture and research applications. Climatic data from Mexico's Sistema de Información Climatológica (SICLIM—Climatological Information System (Velázquez-Alvarez and Balancan-S, 2000) and Eric II (Quintas, 2000)) was integrated into INIFAP's climatic database. Weather stations that are part of this and other climatological networks were georeferenced based on mapping standards defined by Mexico's National Institute for Geography and Statistical Information (INEGI—Instituto Nacional de Geografía e Informática). Quality assurance and quality control issues for integrating climatic geodata from various databases are discussed in the state-level geospatial climate reports produced by INIFAP for its national study on agricultural crops production potential (González-Hernández et al., 2005). The geopositioning of most weather stations has acceptable accuracy standards, while few might have spatial accuracies within 100 m (González-Hernández, 2007). For example, data from weather stations whose datum

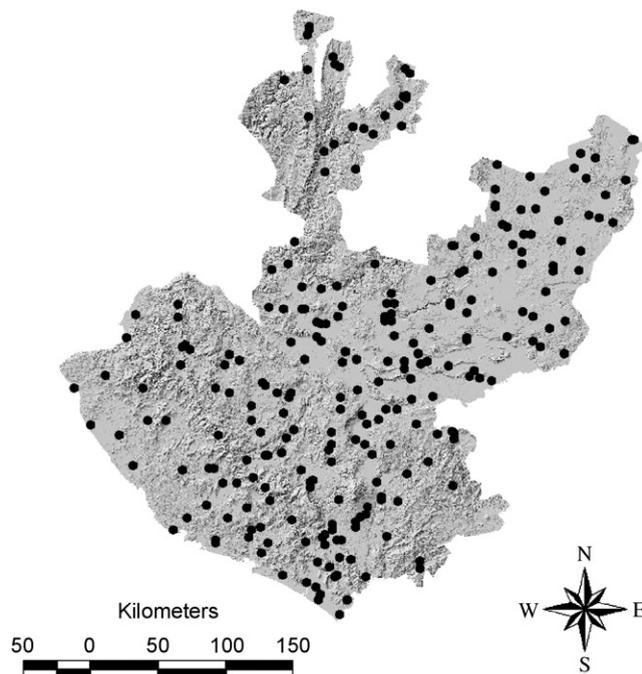


Fig. 1. Location of the weather stations used in modeling climatic data for the states of Jalisco and Colima, Mexico.

was not reported or failed to record data for 30 consecutive years were discarded from INIFAP's integrated climatic database. The climatic data included average monthly temperature ($^{\circ}\text{C}$), precipitation (mm) and evaporation (mm) from 1940 to 1996. Elevations of the weather stations ranged for 4–2505 m.

Temperatures average 19°C in July and 11°C in January, with the lowest recorded average temperature of 1.5°C . Mean annual precipitation averages 73 mm with a dry season from October to May averaging 22 mm from weak cold fronts and a wet season occurring from June to September averaging 162 mm from either adiabatic heating or tropical systems (Garcia, 1965). Prevailing westerly winds have established a west to east precipitation gradient across the two states.

2.3. GIS and Landsat data

Spectral information and topographic data were taken from a satellite imagery and Digital Elevation Model (DEM) of the two states. Ten cloud-free Landsat-7 ETM+ images obtained during the months of January through March 2004 were joined together to create a seamless image. The winter imagery allowed us to better differentiate between major vegetation types (i.e., tropical dry forests, pine–oak forests, grasslands, etc.) than imagery from other seasons. It was hypothesized that the variability in climatic conditions is correlated with the spatial variability in vegetation types as reflected in the satellite imagery. The image was normalized (Hall et al., 1991) to account for differences among scenes and used the normalized image for model development. The satellite imagery consisted of nine spectral bands (spectral bands 1–5, 6L (low gain), 6H (high gain, see USGS-EROS Data Center web site for more information), 7 and 8). Spectral bands 6L and 6H were thermal bands (57 m resolution), while band 8 was a panchromatic image (15 m resolution). These latter three bands were resampled to a 30 m spatial resolution using nearest neighbor techniques (Muukkonena and Heiskanenb, 2005). Nearest neighbor resampling was selected due to quicker computer processing time as compared to other interpolation methods. In addition, nearest neighbor interpolation better maintains original reflectance values while providing sufficient accuracy and reduced potential introduction of unwanted geometric distortions in areas with no ground control points to provide precise control (Muukkonena and Heiskanenb, 2005).

The DEM was obtained from the National Elevation Dataset (NED) as a seamless ArcInfo (ESRI, 1995) grid at a 90 m resolution (U.S. Geological Survey (USGS), Gesch et al., 2002). The DEM was resampled to a 30 m spatial resolution using bilinear techniques (Edenius et al., 2003), producing a more continuous surface reflecting gradual changes in elevation at a 30 m spatial resolution. GIS grids of elevation, slope and aspect were derived from the DEM using ArcView[®] (ESRI, 1998). In addition, a grid of distances from the Pacific Ocean, which borders the western part of the two states, was derived. Values for all grid layers of information were derived for each weather station using Avenue (ESRI, 1998) code.

2.4. Climate modeling

Modeling of the monthly climatic data was accomplished in two stages using procedures developed by Reich et al. (2004). In the first stage, multiple regression analysis was used to describe the coarse-scale variability in the monthly climatic data as a function of elevation, slope, aspect, distance from the coastline, and the Landsat-7 ETM+ bands. For each component of the monthly climatic data, a stepwise procedure was used to identify the best subset of independent variables to include in regression models. Semi-variograms were used to evaluate spatial dependencies among residuals from the various models. If the residuals exhibited spatial dependencies, a generalized least squares model was used to estimate the regression coefficients associated with the regression models (Reich and Davis, 1998). In the second stage, a tree-based stratification design was used to enhance the estimation process of the small-scale spatial variability. With this approach, sample units (i.e., pixel of a satellite image) are classified with respect to predictions of error attributes into homogeneous classes, and the classes are then used as strata in the stratified analysis. Independent variables considered in the binary regression trees (Breiman et al., 1984) included elevation, slope, aspect and Landsat-7 ETM+ bands. A decision rule was used to identify a tree size that minimized the error in estimating the variance of the mean response and prediction uncertainties at new spatial locations.

The effectiveness of the final models were evaluated using a goodness-of-prediction statistic (G) (Agterberg, 1984; Guisan and Zimmermann, 2000; Kravchenko and Bullock, 1999; Schloeder et al., 2001). The G -value

measures how effective a prediction might be relative to that which could have been derived using the sample mean (Agterberg, 1984). A G -value equal to 1 indicates perfect prediction, a positive value indicates a more reliable model than if one had used the sample mean, and a negative value indicates a less reliable model than if one had used the sample mean.

Grids of monthly climatic data were generated for the best fitting regression models. Similarly, grids representing the error associated with each regression model were generated by passing each grid for the appropriate independent variable through the regression trees. The final predicted surfaces were obtained from the sum of the two grids.

2.5. Model evaluation

A 10-fold cross-validation was used to estimate the prediction error for each monthly climate model (Efron and Tibshirani, 1993; Stone, 1974). Data were split into $K = 10$ parts consisting of approximately 25 weather stations and models were fitted to the remaining $K - 1 = 9$ parts of the data. The fitted model was used to predict the part of the data removed from the modeling process. This process was repeated 10 times so that each weather station was excluded from the model fitting step and its response predicted. Prediction errors were obtained from the predicted minus actual values.

Various measures of prediction error were calculated to evaluate the effectiveness of the models. Prediction bias (Williams, 1997) was calculated for each validation data set as a percentage of the true value and accuracy (Kravchenko and Bullock, 1999; Schloeder et al., 2001) was measured by the mean absolute error (MAE), which is a measure of the sum of absolute residuals (i.e., actual minus predicted) and the root mean squared error (RMSE). Small MAE values indicate a model with few errors, while small values of RMSE indicate more accurate predictions on a point-by-point basis (Schloeder et al., 2001).

Estimation uncertainty in the models (Reich et al., 2004) was calculated as the estimation error variance (EEV), $\hat{\sigma}_i^{2*}$ for each observation in the data set. The consistency between the EEV and the observed estimation errors (i.e., true errors) was calculated using the standard mean squared error (SMSE) (Hevesi et al., 1992). EEVs were assumed consistent with true errors if the SMSE fell within the interval $[1 \pm 2(2/n)^{1/2}]$ (Hevesi et al., 1992). Paired t -tests ($\alpha = 0.05$) were used to test for differences between the mean estimation errors and zero. The EEVs were also used to construct 95% confidence intervals around individual estimates. Coverage rates were calculated as the proportion of individual confidence intervals that contained the true value.

2.6. Climate zones

Climatic zones were developed using measures similar to the standardized precipitation index (SPI) developed by Mckee et al. (1993) to characterize variability in precipitation patterns as a standardized departure from some probability distribution. In the simplest case, SPI can be thought of as a z -score associated with a standard normal distribution (Gidding et al., 2005). A value less than zero would indicate that the precipitation is less than normal, while a positive value would indicate that the precipitation is greater than normal. If the data are highly skewed it is recommended to transform the data using either a Gamma, incomplete Beta or Pearson type III distribution when calculating SPI (Gidding et al., 2005; Guttman, 1999).

The first step in defining climate zones was to calculate the cumulative sum of average monthly temperature (T_i), precipitation (P_i) and evaporation (E_i) over the 12 months:

$$\text{TSUM} = \sum_{i=1}^{12} T_i, \quad \text{PSUM} = \sum_{i=1}^{12} P_i, \quad \text{ESUM} = \sum_{i=1}^{12} E_i.$$

Estimates of moisture surplus/deficit (PESUM) were obtained by subtracting ESUM of evaporation from PSUM of precipitation. Positive values indicate a moisture surplus while negative values indicate a moisture deficit.

The distributions of TSUM and PESUM were unimodal and somewhat symmetrical in shape so no transformation was performed in calculating SPI values. The mean and standard deviation of the GIS grids of TSUM and PESUM were calculated and used to compute z -scores for individual grid cells. Using ERDAS-IMAGINE software, a combination of unsupervised and supervised classification procedures was used to

aggregate individual pixels into temperature zones and precipitation–evaporation zones. In the first step cluster analysis was used to group like pixels into zones with similar characteristics. This initial classification was used as a guide in a supervised zoning assignment based on broad vegetation types (tropical, temperate and semi-arid forests). The variability associated with the cumulative distributions, was used to define three temperature zones and four precipitation–evaporation zones. Temperature zones were assigned a value of 10, 20 or 30, and the precipitation–evaporation zones 1, 2, 3 or 4. The temperature zone grid was joined with the precipitation–evaporation grid to produce a grid of the final climate zones. This resulted in 12 unique climate zones that were potentially identified.

2.7. Modeling species distribution

The forests in the states of Jalisco and Colima are some of the most diverse in all of Mexico (Table 1). In Jalisco alone, there are 538 known tree species. The locals use the trees for a variety of purposes—artisan (57 species), drink and liqueur (12 species), construction (42 species), medical (321 species), honey (31 species), pesticides (32 species), domestic use (51 species), landscaping (85 species), fiber (7 species), forage (174 species), industrial timber production (30 species) and veterinarian uses (42 species) (SEMARNAT, 2006). A lot of information is known about the tree species in Jalisco and Colima, and generally where they occur. What is not known is the extent of the distribution of individual tree species in the two states.

To demonstrate the application of the climatic models, logistic regression models were developed to investigate the presence–absence of *Caesalpinia eriostachis* Nee, *Lysiloma acapulcensis* (Kunth) Benth., *Pinus oocarpa* Schiede ex-Schltld. and *Quercus magnoliifolia* Nee as a function of topographic and climatic data. *C. eriostachis* is one of several species that forms the upper forest canopy of the dry tropical forests. *L. acapulcensis* is a temperate deciduous hardwood commonly found in association with pine and oak forests. Stems and branches are used for firewood and in rural construction (SEMARNAT, 2006). The leaves are used for fodder for cattle, while the seeds are used for medicinal purposes. *P. oocarpa* is a species common to the temperate climatic zone. The wood is used for firewood and coal production. The resin is used to produce turpentine as well as for medicinal purposes. *Q. magnoliifolia* is a temperate deciduous hardwood species. The stems are used to make tool handles and fence posts. It is also used as firewood and in the manufacture of coal. The acorns are edible and are used in the preparation of tortillas.

Presence–absence data for the four species came from a state-wide inventory consisting of 1442 permanent plots established in 2006 using a two-way nested design (Reich et al., 2006). For each species model parameters were estimated using maximum likelihood methods. Model and predictor significance were obtained from Wald test statistic assuming a Chi-square distribution with one degree of freedom (Harrell, 2001). Full models were reduced using backward elimination. The relative fit of each model to the data was evaluated using Akaike Information Criteria (AIC, Akaike, 1973). For a given species the model with the lowest AIC value is that which is most likely after penalization for the number of parameters. AIC values allow discrimination of the relative performance among competing models, however, this may not be the most important criteria used in selecting a model. Receiver operating curves (ROC) were used to assess the predictive accuracy of the binary models (Boyce et al., 2002). A 10-fold cross-validation was used to assess the predictive performance of the binary models. Binary maps based on the probability of the inverse logistic function of the linear predictors were developed at a 30 m spatial resolution.

Table 1
Trees species diversity in the state of Jalisco, Mexico, by climatic zone

Climatic zone	Number of tree species counted on sample plots	Number of trees counted on sample plots	Diversity index (Shannon–Weaver index)
Semi-arid	88	1185	3.40
Temperate	165	3321	3.67
Tropical	214	2963	4.48

The data come from a state-wide inventory consisting of 1442, 30 m × 30 m permanent plots allocated using a stratified design (Reich et al., 2006).

To understand the spatial pattern associated with the binary maps developed for the four tree species, a simple product kernel was used to estimate the intensity, or the number of events (presence of a tree species in a 30 m × 30 m pixel) per unit area in a circle with a radius of 5000 m. The decision to use a radius of 5000 m was based on a previous study of the spatial dependencies of *Armillaria* root disease in the Black Hills (Kallas et al., 2003). The kernel function in two dimensions is defined as

$$\hat{\lambda}_{5000}(s) = \sum_{i=1}^N \rho_{5000}(u_i) * \rho_{5000}(u_i),$$

where

$$\rho_{5000}(u) = \begin{cases} 0.75[1 - (u/5000)^2]/5000, & -5000 \leq u \leq 5000, \\ 0, & \text{otherwise} \end{cases}$$

is a probability density function symmetric about the origin, and u is the distance from the center of a pixel indicating the presence of a given species to some arbitrary location, s (Cressie, 1991). This information was converted to a GIS layer and contours generated to aid in interpretation.

3. Results

3.1. Climate modeling

The observed variability in average monthly temperature, precipitation and evaporation data used in developing the models are presented in Fig. 2. Although the majority of the sample distributions were skewed, residuals plots and plots of predicted vs. observed climatic values did not indicate any trends to suggest a systematic bias in any of the models.

The climatic data were linearly correlated with various topographic and Landsat-7 ETM+ bands and these relationships varied from month to month and the variable being modeled. For example, in the evaporation models, distance from the coastline was the most important variable while in the temperature models, elevation and Landsat band1 were the two most important variables. All combinations of topographic and Landsat-7 ETM+ bands were used in regression trees to describe the error in one or more of the regression models. The tree sizes selected to minimize the error in estimating the variance of the mean response and prediction uncertainties at new spatial locations ranged from 28 to 40 terminal nodes.

The overall contribution of the models (Tables 2–4) in describing temperature, precipitation and evaporation varied with the model. The regression models for temperature alone explained 32% (May) to 72% (September) of the observed variability, while in the precipitation and evaporation models, the regression models only explained from 7% (November evaporation) to 43% (September precipitation). The binary regression trees accounted for an additional 11% (September temperature) to 58% (July precipitation) of the unexplained variability in the regression models. Overall model performance for temperature ranged from a low of 0.57 for May to a high of 0.85 for October with a median of 0.81. The median prediction performance for precipitation models was 0.67 with a low of 0.45 for January to a high of 0.80 for September. The evaporation models had a median of 0.72 and ranged from 0.49 for October to 0.77 for August.

3.2. Model evaluation

Prediction bias was nominal (Tables 2–4) for all models. The models generally underestimated the left tail of the distributions. This is especially evident for the precipitation models for July and August and the evaporation models for April and May (Fig. 2). The mean estimation errors did not differ significantly from zero (p -value ≥ 0.05). The MAE was smaller than the RMSE for all models and indicated that models are more accurate in predicting regional or global means than on a point-by-point basis.

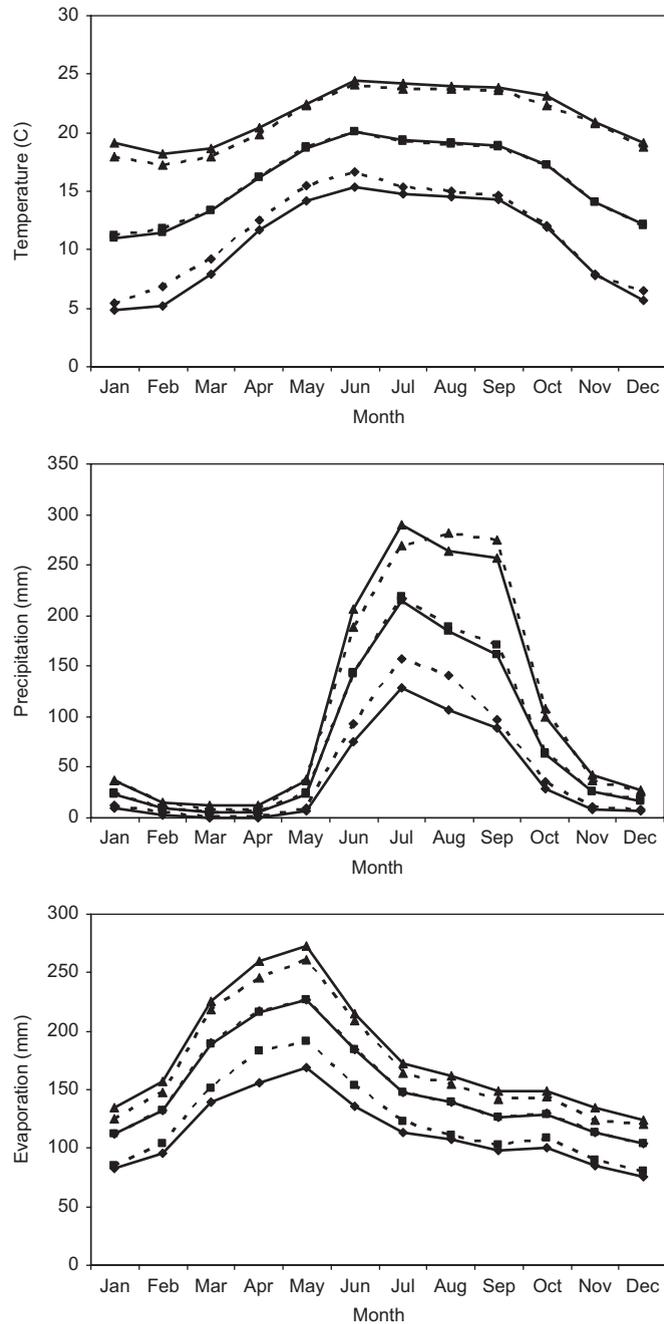


Fig. 2. Observed (solid lines) and predicted (dashed lines) variability in the distribution of average monthly temperature (°C), precipitation (mm) and evaporation (mm) for the states of Jalisco and Colima, Mexico. In a given graph, the upper and lower pair of lines (solid and dashed) represent the 0.05 and 0.90 quantile values of the distribution, respectively, while the pair of lines in the middle represent average values.

SMSE results (Tables 2–4) showed that the computed EEVs consistently estimated the true errors for the models, as they generally fell within the interval [0.809–1.191] (Hevesi et al., 1992). The SMSEs were fairly consistent for all months and all models suggesting that it would be possible to empirically assess estimates of uncertainty for new observations. The 0.95 confidence coverage rates ranged from a low of 0.92 to a high of

Table 2

Summary statistics of the predictive performance of the temperature (°C) models developed for the states of Jalisco and Colima, Mexico based on a ten-fold cross-validation

Month	Statistic ^a				
	G-statistic	Bias (°C)	RMSE (°C)	SMSE	0.95 Confidence coverage
January	0.810	0.2	3.5	0.92	0.97
February	0.736	0.0	3.8	0.98	0.99
March	0.673	0.0	3.8	0.93	0.98
April	0.648	0.0	3.8	0.99	0.98
May	0.574	−0.1	3.6	0.84	0.98
June	0.772	0.0	2.6	1.01	0.97
July	0.835	0.0	2.0	0.97	0.98
August	0.841	0.0	2.0	0.81*	0.98
September	0.828	0.0	2.0	1.05	0.96
October	0.854	0.0	2.5	0.68*	0.97
November	0.816	0.1	3.2	1.00	0.96
December	0.836	0.1	3.1	1.01	0.97

^aG-statistic used to measure the effectiveness of the predictions; RMSE, root mean squared error; SMSE, standardized mean square error.

*SMSE is significantly different from 1 at the 0.05 level of significance.

Table 3

Summary statistics of the predictive performance of the precipitation (mm) models developed for the states of Jalisco and Colima, Mexico, based on a 10-fold cross-validation

Month	Statistic ^a				
	G-statistic	Bias (mm)	RMSE (mm)	SMSE	0.95 Confidence coverage
January	0.450	0.7	19.1	2.76*	0.94
February	0.674	0.0	6.1	0.75*	0.96
March	0.558	0.1	5.7	1.05	0.95
April	0.530	−0.2	6.1	1.00	0.95
May	0.675	−0.2	11.9	0.66*	0.98
June	0.680	2.1	44.2	0.68*	0.97
July	0.727	−5.3	67.2	0.79*	0.96
August	0.732	−3.3	69.7	0.94	0.94
September	0.802	−9.5	72.6	0.99	0.92
October	0.769	−0.9	31.2	1.04	0.92
November	0.470	−0.3	21.1	0.99	0.97
December	0.511	−0.4	11.1	0.78*	0.96

^aG-statistic used to measure the how effective the prediction; RMSE, root mean squared error; SMSE, standardized mean square error.

*SMSE is significantly different from 1 at the 0.05 level of significance.

0.98. Deviations from the expected confidence coverage rates are due the errors in estimating the uncertainty associated with estimates.

3.3. Climate zones

Fig. 3 shows the temperature and precipitation–evaporation zones for the states of Jalisco and Colima. The criteria used to define the temperature and precipitation–evaporation zones are summarized in Tables 5 and 6. When the temperature zones were combined with the precipitation–evaporation zones this yielded 12 unique

Table 4

Summary statistics of the predictive performance of the evaporation (mm) models developed for the states of Jalisco and Colima, Mexico, based on a 10-fold cross-validation

Month	Statistic ^a				
	G-statistic	Bias (mm)	RMSE (mm)	SMSE	0.95 Confidence coverage
January	0.628	1.5	20.7	0.57*	0.98
February	0.714	−0.1	23.7	1.11	0.92
March	0.685	−0.5	32.2	0.89	0.96
April	0.727	−1.9	34.0	0.73*	0.97
May	0.756	0.6	47.9	0.76*	0.97
June	0.734	−1.9	27.3	0.73*	0.96
July	0.750	−1.1	20.3	0.97	0.94
August	0.770	0.4	18.1	0.93	0.96
September	0.737	−0.2	17.2	1.00	0.94
October	0.495	−0.7	25.1	1.17	0.97
November	0.599	0.4	19.1	1.13	0.93
December	0.644	0.1	20.4	1.28*	0.92

^aG-statistic used to measure the how effective the prediction; RMSE, root mean squared error; SMSE, standardized mean square error.

*SMSE is significantly different from 1 at the 0.05 level of significance.

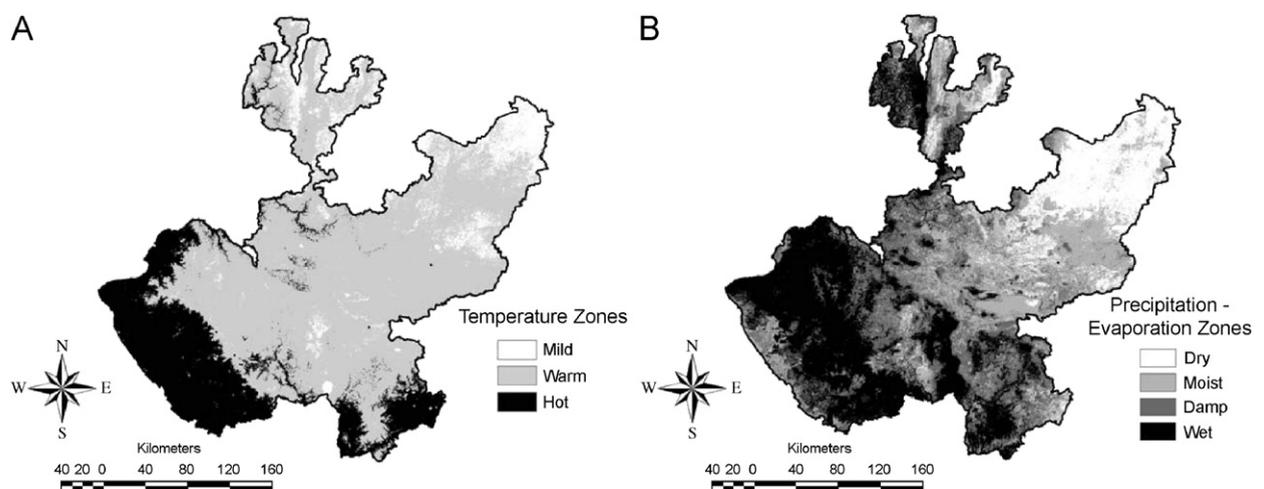


Fig. 3. Derived temperature (A) and precipitation–evaporation (B) zones for the states of Jalisco and Colima, Mexico.

ecological regions defined by climate, topography and vegetation. These regions coincide in general with those adopted to describe the vegetation and ecology in Mexico (Rzedowski, 1978). The climatic zones define three broad ecological regions. The first is the *sub-humid tropical zone*, which is located along the Pacific coast. The region is characterized by hot temperatures, monsoon rains during the summer months and an annual dry period of 5–9 months. Tropical dry forests dominate this region. At higher elevations is the *sub-humid temperate zone*, which covers the greatest portion of the two states. Pine–oak and mixed deciduous hardwood forests dominate this region. This region gradually changes to an *arid and semi-arid zone*, which has a low annual precipitation with 6–8 dry months. The semi-arid region is dominated by mesquite-grasslands while xerophytic scrublands dominate the arid region.

The spatial variation in vegetation is consistent with the temperature and precipitation–evaporation zones (Fig. 3). The increase in tree species diversity (Table 1) from east to west is likely a result of increased

Table 5

Description and criteria used to define temperature zones in the states of Jalisco and Colima, Mexico, based on the cumulative sum of temperatures over the 12 months

Zone	Description	Range (°C)	Definition
10	Mild	24.6–150.3	<−1 Standard deviation
20	Warm	150.3–229.0	−1 Standard deviation to +1 standard deviation
30	Hot	229.0–319.9	> +1 Standard deviation

Table 6

Description and criteria used to define precipitation–evaporation zones in the states of Jalisco and Colima, Mexico, based on the cumulative sum over the 12 months

Zone	Description	Range (mm)	Definition
1	Dry	−2894.0–2151.9	<−1 Standard deviation
2	Moist	−2151.9–1780.8	−1 Standard deviation to 0 standard deviation
3	Damp	−1780.8–1409.6	0 Standard deviation to +1 standard deviation
4	Wet	−1409.6–102.6	> +1 Standard deviation

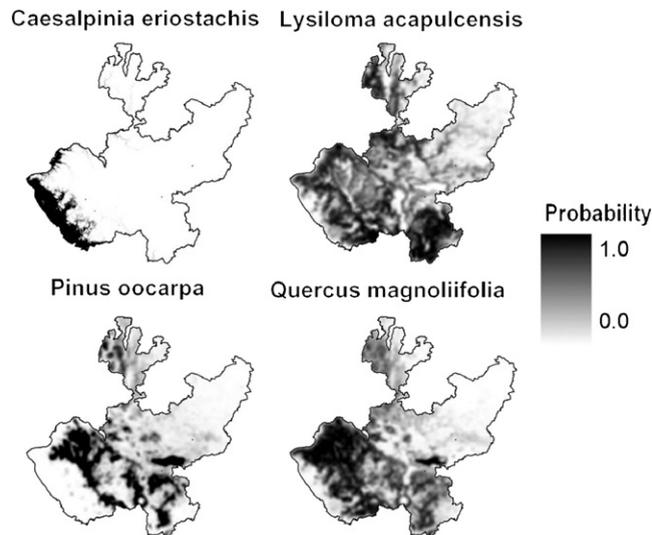


Fig. 4. Spatial distribution of the estimated probability of observing a selected tree species in the states of Jalisco and Colima, Mexico. Probabilities were computed using a kernel density estimator with a bandwidth of 5 km.

precipitation (Gentry, 1995). Rainfall and temperature have long been known as chief factors that determine tropical vegetation (Holdridge, 1947).

3.4. Species distributions

Enough data were available from the state-wide inventory to try and generate distribution maps for four tree species (*C. eriostachis* (7% of sample plots), *L. acapulcensis* (10% of sample plots), *P. oocarpa* (13% of sample plots) and *Q. magnoliifolia* (7% of sample plots)) (Fig. 4). Different combinations of topographic and climatic data were used in logistic regression models to describe the presence–absence of the selected tree species on the sample plots. Temperature variables were the most frequently selected. Summary statistics describing the accuracy of the models are given in Table 7. These models were used to predict the presence or absence of each tree species. These surfaces were then subjected to a kernel estimator to generate maps depicting the probability of observing a particular tree species (Fig. 4).

Table 7

Statistics summarizing the accuracy of species distribution models for the states of Jalisco and Colima, Mexico

Species	Overall accuracy	Errors of omission		Errors of commission	
		Absence	Presence	Absence	Presence
<i>Caesalpinia eriostachis</i>	0.66	0.35	0.17	0.02	0.84
<i>Lysiloma acapulcensis</i>	0.62	0.39	0.12	0.02	0.85
<i>Pinus oocarpa</i>	0.84	0.17	0.16	0.01	0.82
<i>Quercus magnoliifolia</i>	0.77	0.19	0.23	0.02	0.81

The spatial models for the four tree species show that the selected tree species had different distributions across the landscape. Landscape patterns differ in extent and shape. *C. eriostachis* has the most limited distribution while *L. acapulcensis* had the widest distribution. *P. oocarpa* distribution is limited to higher elevation temperate forests while *Q. magnoliifolia* is found in both the tropical and temperate regions.

Accuracy assessments indicate the models overestimated species occurrences (Table 7). The state-wide inventory was not designed as a vegetation survey and given the diversity of the forests, the low sampling intensity associated with the sampling design, it is not surprising that the models overestimated species occurrences. If such models are used for resource assessments, planners must be aware of the implications of the false-positives and false-negatives errors and how they may impact management decisions.

4. Discussion and conclusions

This study presents a description of a framework for modeling selected climatic data at a fine spatial resolution and for characterizing the error inherent in the technique. Most climate data are modeled at a very coarse spatial resolution for the purpose of describing global trends in the climate (Cohen, 1990; Gates, 1985; Lamb, 1987). However, the influence of climate on the distribution of certain vegetation types, the survival rates of plantations, and the presence/absence of rare and endangered species occur at a finer spatial resolution. Characterizing and quantifying the influence climate has on the distribution of vegetation requires detailed information on the spatial distribution of climate data at a fine enough spatial resolution to accurately describe this dynamic interaction. This study developed a series of models describing the spatial distribution of average monthly temperature (°C), precipitation (mm), and evaporation (mm) in the states of Jalisco and Colima to a 30 m resolution. The models provide unbiased estimates of the climate data as well as estimates of the prediction variance associated with individual estimates.

Derived models described 57–85% of the spatial variability in the temperature, 45–80% in precipitation and 49–77% in evaporation. The poor performance of some models to describe various climate components may be due, in part, to some of the extreme weather conditions reported by some weather stations that made it difficult to adequately model all data. This in turn lowered the overall accuracy of the model.

The ability to calculate estimation uncertainties allowed us to develop GIS layers showing the computed estimation errors as well as place confidence intervals around estimates. It was assumed that the estimation uncertainties are at their lower limit because the data used in this study were assumed error free. Other sources of errors that could have influenced the performance of the models included the sparseness of the weather stations, errors in the recording the climatic data, and registrations errors in the location of weather stations.

To demonstrate the application of these models maps were generated to predict the probability of occurrence of four selected tree species (*C. eriostachis*, *L. acapulcensis*, *P. oocarpa* and *Q. magnoliifolia*) and are based on the occurrence and distribution of the tree species, climate and topography. These maps do not necessarily indicate where these tree species currently occur, but rather where they have a predicted likelihood of being found. Since the models are based primarily on climatic data this may have resulted in a wider range of land cover types being deemed suitable. The climatic data may be emphasizing the fundamental niche (Kearney and Porter, 2004) of the tree species rather than their realized niche. The inclusion of additional environmental and habitat-related variables could improve the models such that they could provide insight into factors limiting species distributions and how they respond to climatic change.

These species distribution models rely heavily on the resolution and the reliability of the climatic models used as input. The climatic models provided detailed estimates of the distribution of temperature, precipitation and evaporation at a fine scale (30 m resolution) with relatively high accuracy. These models can, in general, produce greater accuracy and spatial resolution than what is currently available for the region.

Our initial interest in developing and evaluating these models was aimed at characterizing and quantifying the vegetation in this region. Because of the diversity of the vegetation in the two states it is important that accurate vegetation maps be developed. Typically, vegetation maps are generally developed using remotely sensed imagery (i.e., Landsat-7 ETM+). There are some vegetation types (*Juniperus* forests) in the eastern part of Jalisco (semi-arid region) that have similar spectral properties to forests in the temperate region making it difficult to accurately model certain vegetation types. The natural vegetation that occurs in these two states is controlled to a large degree by the climate. Being able to accurately model the climate one should be able to develop more accurate and reliable vegetation maps for the two states.

The climatic zones developed using the climatic data are not unique. They were developed primarily to study other climate-related phenomena. It would be possible to derive much smaller, more detailed zones, or to derive larger zones by combining several of the smaller zones. The climatic zones provide a unique perspective on the behavior of temperature, precipitation and evaporation in the area. The derived climatic models could be used to provide a basis for the analysis of climatic impacts on the distribution of selected plant species, identify suitable species for reforestation efforts in selected areas and the detection of invasive species, to name just a few applications. Determining the species–environment relationship is an important issue in ecology (Guisan and Zimmermann, 2000) and the climatic models developed in this study provide new opportunities for future research in this area.

Acknowledgments

The authors would like to thank the support of the Mexican state of Jalisco's Secretary of Rural Development, FIPRODEFO (Fideicomiso para la Administracion del Programa de Desarrollo Forestal del Estado de Jalisco); Jalisco's Agriculture Advisory Board, the USDA Forest Service, Rocky Mountain Research Station, Colorado State University, and the Universidad de Colima, Mexico. Each of these institutions and organizations provided the means and human resources for making this project possible.

References

- Agterberg, F.P., 1984. Trend surface analysis. In: Gaile, G.L., Willmott, C.J. (Eds.), *Spatial Statistics and Models*. Reidel, Dordrecht, The Netherlands, pp. 147–171.
- Akaike, H., 1973. Information theory and extension of the maximum likelihood principle. In: Petrov, N., Csaki, F. (Eds.), *Second International Symposium on Information Theory*. Hungarian Academy Sciences, Budapest, Hungary, pp. 268–281 (Reprinted in: Kotz, S., Johnson, N. (Eds.), *Breakthroughs in Statistics*, vol.1. Springer, New York, USA, 1992, pp. 610–624).
- Allen, T.F.H., Hoekstra, T.W., 1992. *Toward a Unified Ecology*. Columbia University Press, Columbia, 384p.
- Bailey, R.G., 1996. *Ecosystem Geography*. Springer, New York, 204p.
- Bailey, R.G., 2002. *Ecoregions-based Design for Sustainability*. Springer, New York, 223p.
- Bonan, G.B., 2002. *Ecological Climatology: Concepts and Applications*. Cambridge University Press, New York, 610p.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluation resource selection functions. *Ecological Modeling* 157, 281–300.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, I.J., 1984. *Classification and Regression Trees*. Wadsworth, Pacific Grove.
- Burrough, P.A., 1986. *Principles of Geographic Information Systems for Land Assessment*. Oxford University press, New York, 193p.
- Challenger A., 1998. *Utilizacion y conservacion de los ecosistemas terrestres de Mexico. Pasada, presente y futuro*. Comision Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico, D.F., 847p.
- Cohen, S.J., 1990. Bringing the global warning issue closer to home: the challenge of regional impact studies. *Bulletin of the American Meteorological Society* 71, 520–526.
- Cressie, N., 1991. *Statistics for Spatial Data*. Wiley, New York, 900p.
- de Sherbinin, A., 2002. *A CIESIN thematic guide to land-use and land-cover change (LUCC)*. Center for International Earth Science Information Network (CIESIN) Columbia University Palisades, NY, USA (Available online at: <http://sedac.ciesin.columbia.edu/guides/lu/CIESIN_LUCC_TG.pdf>).
- Edenius, L., Vencatasawmy, C.P., Sandstrom, P., Dahlberg, U., 2003. Combining satellite imagery and ancillary data to map snowbed vegetation important to Reindeer *Rangifer tarandus*. *Arctic, Antarctic and Alpine Research* 35, 150–157.

- Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York, 436p.
- Englund, E.J., 1990. A variance of geostatisticians. *Journal of Mathematical Geology* 22, 417–455.
- ESRI, 1995. ARC/INFO[®] Software and On-line Help Manual. Environmental Research Institute, Inc., Redlands, CA.
- ESRI, 1998. ArcView[®] 3.1. Environmental Research Institute, Inc., Redlands, CA.
- García, E., 1965. Distribucion de la precipitacion en la Rebulica Mexicana. *Publicacao do Instituto de Geografia UNAM* 1, 173–191.
- Gates, W.L., 1985. The use of general circulation models in the analysis of the ecosystem impacts of climatic change. *Climatic Change* 7, 267–284.
- Gentry, A.H., 1995. Diversity and floristic composition of neotropical dry forests. In: Bullock, S.H., Mooney, H.A., Medina, E. (Eds.), *Seasonally Dry Forests*. Cambridge University Press, Cambridge, UK, pp. 146–194.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., Tyler, D., 2002. The national elevation dataset. *Photogrammetric Engineering and Remote Sensing* 68, 5–32.
- Gidding, L., Soto, M., Rutherford, B.M., Maarouf, A., 2005. Standardized precipitation index zones for Mexico. *Atmosfera* 18, 33–56.
- González-Hernández, A., 2007. Personal communication. Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP), Coyoacán, México, D.F., México.
- González-Hernández, A., Cruz-Bello, G.M., Moreno-Sánchez, F., Velasco-Bautista, E., 2005. Comparación de distribuciones probabilísticas para generar información climática en análisis de aptitud del terreno. *Ciencia Forestal*, vol. 30 (Julio–Diciembre), Num. 98. Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP), Coyoacán, México, D.F., México.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 47–186.
- Guttman, N.B., 1999. Accepting the standardized precipitation index: a calculation index. *Journal of the American Water Resources Association* 35, 311–322.
- Hall, F.G., Strelbel, D.E., Nickeson, J.E., Gotz, S.J., 1991. Radiometric reflection: toward a common radiometric response among multidata multisensor images. *Remote Sensing of Environment* 35, 11–27.
- Harrell, F.E., 2001. *Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics. Springer, New York, 568p.
- Heine, G.W., 1986. A controlled study of some two-dimensional interpolation method. *COGS Computer Contributions* 3, 60–67.
- Hevesi, J.A., Istok, J.D., Flint, A.L., 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. *Journal of Applied Meteorology* 31, 661–676.
- Holdridge, L.R., 1947. Determination of the world plant formations from simple climatic data. *Science* 105, 267–368.
- INEGI, 2001. Estudio hidrológico del estado de Jalisco. Instituto Nacional de Estadística Geografía e Informática (INEGI), segunda edición, Mexico, D.F., Mexico.
- IPCC, 1994. IPCC technical guidelines for assessing climatic change impacts and adaptations. In: Carter, T.R., Parry, M.L., Harasawa, H., Nishioka, S., WMO/UNEP (Eds.), Working Group II. CGER-IO15-'94. University College London, UK and Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan, 59p.
- Kallas, M.A., Reich, R.M., Jacobi, W.R., Lundquist, J.E., 2003. Modeling the probability of observing *Armillaria* root disease in the Black Hills. *Forest Pathology* 33, 241–252.
- Kearney, M., Porter, W.P., 2004. Mapping the fundamental niche: physiology, climate and the distribution of a nocturnal lizard. *Ecology* 85, 3119–3131.
- Kravchenko, A., Bullock, D.G., 1999. A comparative study of interpolation methods for mapping soil properties. *Agronomy Journal* 91, 393–400.
- Lamb, P., 1987. On the development of regional climatic scenarios for policy oriented climatic impact assessments. *Bulletin of the American Meteorological Society* 68, 1116–1123.
- McKee, T.B., Doesken, N.S., Kleist, J., 1993. The relationship of drought frequency and duration of time series. In: Eighth Conference on Applied Climatology, American Meteorological Society, Anaheim, CA, January 17–23, 1993, pp. 179–186.
- Muukkonena, P., Heiskanen, J., 2005. Estimating biomass for boreal forests using ASTER satellite data combined with standwise inventory data. *Remote Sensing of Environment* 99, 434–447.
- Nalder, I.A., Wein, R.W., 1998. Spatial interpolation of climatic normals; test of a new method in a Canadian boreal forest. *Agricultural and Forest Meteorology* 92, 211–225.
- Nixon, K.C., 1993. El genero *Quercus* em Mexico. In: Ramamoorthy, T.P., Bye, R., Lot, A., Fa, J. (Eds.), *Diversidad Biologica de Mexico. Origenes y Distribucion*. Instituto de Biología, UNAM, Mexico.
- Oliver, M.A., 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographic Information Systems* 4, 313–332.
- Quintas, I., 2000. ERIC II-Extractor rápido de información climatológica (VERSION 2.0). Instituto Mexicano de Tecnología del Agua (IMTA), Morelos, México <<http://www.imta.gob.mx>>.
- Reich, R.M., Davis, R.A., 1998. On-line Spatial Library for the S-PLUS[®] Statistical Software Package. Colorado State University, Fort Collins, Colorado.
- Reich, R.M., Lundquist, J.E., Bravo, V.A., 2004. Spatial models for estimating fuel loads in the Black Hills, South Dakota, USA. *International Journal of Wildland Fire* 13, 119–129.
- Reich, R.M., Aguirre-Bravo, C., Schreuder, H., 2006. Statistical strategy for monitoring and assessment of land resources in the state of Jalisco, Mexico. *Reporte Técnico* 5, FIPRODEFO, Guadalajara, Mexico, 61p.
- Royle, A.G., Clausen, F.L., Frederiksen, P., 1981. Practical universal kriging and automatic contouring. *Geoprocessing* 1, 337–394.
- Rzedowski, J., 1978. *Vegetacion de Mexico*. Editorial Limusa, Mexico, D.F., Mexico.
- Schloeder, C.A., Zimmermann, N.E., Jacobs, M.J., 2001. Comparison of methods for interpolating soil properties using limited data. *American Society of Soil Science Journal* 65, 470–479.

- Schreuder, H.T., Williams, M.S., Aguirre-Bravo, C., Patterson, P.L., 2003. Statistical strategy for inventorying and monitoring the ecosystem resources of the Mexican states of Jalisco and Colima at multiple scales and resolution levels. Report RMRS-GTR-107, USDA Forest Service, Fort Collins, Colorado, USA.
- Schultz, J., 1995. *The Ecozones of the World: The Ecological Divisions of the Geosphere*. Springer, Berlin, 252p.
- SEMARNAT, 2006. *Especies con usos no maderables en bosques de encino, pino y pino-encino en los estados de Chihuahua, Durango, Jalisco, Michoacán, Guerrero y Oaxaca* <<http://www.semarnat.gob.mx/pfnm>>.
- Stone, M., 1974. Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B* 36, 111–133.
- Velázquez-Alvarez, J., Balanca-S, A., 2000. SICLIM 1.0—Sistema de Información Climatológica. Base de Datos, Período 1921–1990. Instituto Mexicano de Tecnología del Agua, IMTA, Morelos, México <<http://www.imta.gob.mx>>.
- Williams, M.S., 1997. A regression technique accounting for heteroscedastic and asymmetric error. *Journal of Agriculture, Biology and Environmental Statistics* 2, 108–129.