Ecological Archives A022-007-A1

Gerald E. Rehfeldt, Nicholas L. Crookston, Cuauhtémoc Sáenz-Romero, and Elizabeth M. Campbell. 2012. North American vegetation model for land-use planning in a changing climate: a solution to large classification problems. *Ecological Applications* 22:119–141.

Appendix A. Acquisition of data points used in fitting a Random Forests classification tree to point locations within 46 North American biomes.

We used five steps to obtain a sample of locations from the 5707 polygons in the vegetation classification of Brown et al. (1998). The first step was to obtain from each polygon a sample of data points referenced by latitude and longitude. For this, we developed a sampling protocol whereby the number of data points to be selected from each polygon was determined by the area of the polygon. The MAPTOOLS package in R (2004) was used to extract data points according to a series of rules that would extract one data point for each km2 in polygons less than 100 km2, but as polygon size increased, extract fewer data points. Accordingly, the second rule pertained to polygons with 100 < km^2 < 5000, and extracted 100 data points plus an amount equaling ½ of the area exceeding 100 km^2. Rules for larger polygons were approximately equivalent to:

$$N = 8.5A^{0.5} + 303.9A^{0.25}$$

where N is the number of data points and A is the area (km^2) of the polygon (Fig. A1). Of the software options, we used a regular sampling with 20 iterations and imposed an offset of 0.5 to limit sampling polygon edges. Data points from polygons smaller than 10 km^2 were discarded to reduce potential effects of misalignment of the geographically referenced files.

According to this sampling protocol (Fig. A1), a polygon of 100 km^2 would be allocated 100 data points; one of 1000 km^2 would receive 550; and the largest polygon of 5 million km^2 would receive 30,000. The procedure thus provided a strong representation of small polygons while assuring that data points from the largest polygons would not dominate the dataset (Fig. A2).

Next, biomes were culled that were peripheral to our objectives: those from Greenland and from the Caribbean Islands; those from south of 13.9 °N latitude, the southern limits of our climate model; and those designating wetlands. Of the remaining biomes, those of similar vegetation that were geographically disparate were combined. For example, the original biotic communities of Brown et al. (1998) included separate entries for Central American Savanna, Guerrero Savanna, and Campeche – Veracruz Savanna that were combined into a single biome which we call Tropical Savanna. The resulting dataset contained 46 biomes and about 1.6 million point locations, the latitude and longitude of which were exported into a data file, referred to subsequently as biome data points.

In the second step, the elevation of each data point was estimated from the digitized elevations (0.0083° resolution) of GLOBE Task Team (1999). Spline climate surfaces (see Rehfeldt 2006; Saenz-Romero et al. 2010), available at URL: http://forest.moscowfsl.wsu.edu/climate/, could then be used to estimate 18 climate variables relevant to plant geography (see Rehfeldt 2006, Rehfeldt et al. 2006, Saenz-Romero et al. 2010) for each point location. These variables, plus several interactions between temperature and precipitation variables, provided 34 variables of potential use in predicting the occurrence of biomes.

The third step was to use supplementary data in an attempt to reduce errors in the classification of data points from those biomes most affected by the imperfect alignment of geographically referenced files and thereby contributing disproportionately to the errors of fit in previous analyses (Rehfeldt et al. 2006) and in previous iterations of the present analysis:

1. Biome 37: all biome data points were deleted. The lower altitudinal limits of the eastern USA subalpine forests and tundra were obtained from Sabo and Whittaker (1979), Hall (1988), and Burns and Honkala (1990). These limits were verified using FIA plot data for the distributions of *Abies* spp. or *Picea rubens*. Globe Task Team's (1999) DEMs were searched to find locations above the lower limits that ranged from 850 m in the northern mountains and 1700 m in the south. The procedure located about 15,000 locations that were added to the set of biome data points.
2. Western Canada: From within the area bounded by 60 > latitude > 49 and long < -114 in Canada, all biome data points were discarded and were replaced with a sampling from polygons in the digitized map of British Columbia's ecosystems (Meidinger and Pojar 1991). In merging the British Columbia classification system to that of Brown et al. (1998), the Alpine Tundra class was assigned biome 22; the Interior Douglas-fir, Ponderosa Pine, and Sub-Boreal Pine-Spruce classes to biome 43; the Engelmann.
3. Spruce-Subalpine Fir, Montane Spruce, and Sub-Boreal Spruce classes to biome 41; the Boreal Black and White Spruce and Spruce Willow-Birch to biome 50; Coastal Mountain Heather Alpine and Boreal Altai Fescue Alpine to 44; Coastal Western Hemlock to 36; and Bunchgrass to 42. Those data points assigned to biomes 43 or 41 that were south of 50.86° and west of -119.7° were re-assigned to biomes 32 and 21, respectively. These procedures produced about 241,000 observations that were added to the set of biome data points.
4. 4. Mexico: the biomes of concern were 2, 17, and 22 in the Transvolcanic Axis, 20.4 > latitude > 18.2 by -97 > longitude > -103.6. For 22, biome data points were replaced with data points above timberline on seven volcanoes (see Ledig et al. 2010) plus those from GLOBE Task Team (1999) for a similar elevation range; the total was about 900 observations. For 2 and 17, unpublished forest inventory data were used to supplement biome data points, using the presence of *Abies guatemalansis*, *Abies religiosa*, *Pseudotsuga menziesii*, *Pinus hartwegii*, and *Pinus rudis* (considered by some as low altitudinal populations of *P. hartwegii*, Matos 1995) as indicators of biome 2; and *Pinus devoniana* (also known as *P. michoacana*, Sáenz-Romero and Tapia-Olivares 2008), *P. douglasiana*, *P. greggii*, *P. lawsonii*, *P. maximinoi*, *P. oaxacana*, and *P. pringlei* as indicators of biome 17. The range in elevation of species in the two groups was used to cull biome data points whose elevations were flagrantly in error. Altogether, about 450 new observations were added to the biome data set.
5. Western USA: The biomes of concern were 43, 41, and 22 which ordinarily occur as an ascending altitudinal sequence in the Rocky Mountains and Great Basin. For these biomes, all biome data points were discarded for 49 > latitude > 30. For biome 22, upper timberlines within 14 geographic sub-regions (Arno and Hammerly 1984) set the lower altitudinal limits of the biome. DEMs of GLOBE Task Team (1999) were searched to locate about 10,000 data points at elevations higher than the threshold. For biomes 43 and 41, original biome data points were run through the climate profiles (URL: http://forest.moscowfsl.wsu.edu/climate/) of *Pinus ponderosa*, *Pseudotsuga menziesii*, or *Thuja plicata* to determine the suitability of the climate for biome 43; of *Picea engelmannii*, *Abies lasiocarpa*, and *A. lasiocarpa* var. *arizonica* for suitability for biome 41; and of *Pinus edulis* for suitability for biome 8. Timberline altitudes also were used to assort biome data points between biomes 44 and 46.
6. Biome 25: An omission in the classification of Brown et al. (1998) is the Interior Cedar-Hemlock Ecosystem in northwestern USA and adjacent Canada (see Daubenmire and Daubenmire 1968). For Canada, we obtained a sample of data points from the digitized file of Meidinger and Pojar (1991). For the USA, all biome data points for biome 43 were discarded for 49 > latitude > 45.9 and -118.7 < longitude < -113. Original biome 43 polygons within this area were resampled to

obtain approximately 100,000 observations within the polygon delimited above. The data points and their climates were run through the climate profile of *Thuja plicata*. The data points receiving more than 55 % of the votes for the climate being suitable to *Thuja plicata* were re-assigned to biome 25. The procedures produced approximately 24,000 data points for biome 25 from each country.

These steps resulted in a database with about 1.75 million observations spread across 46 biomes.

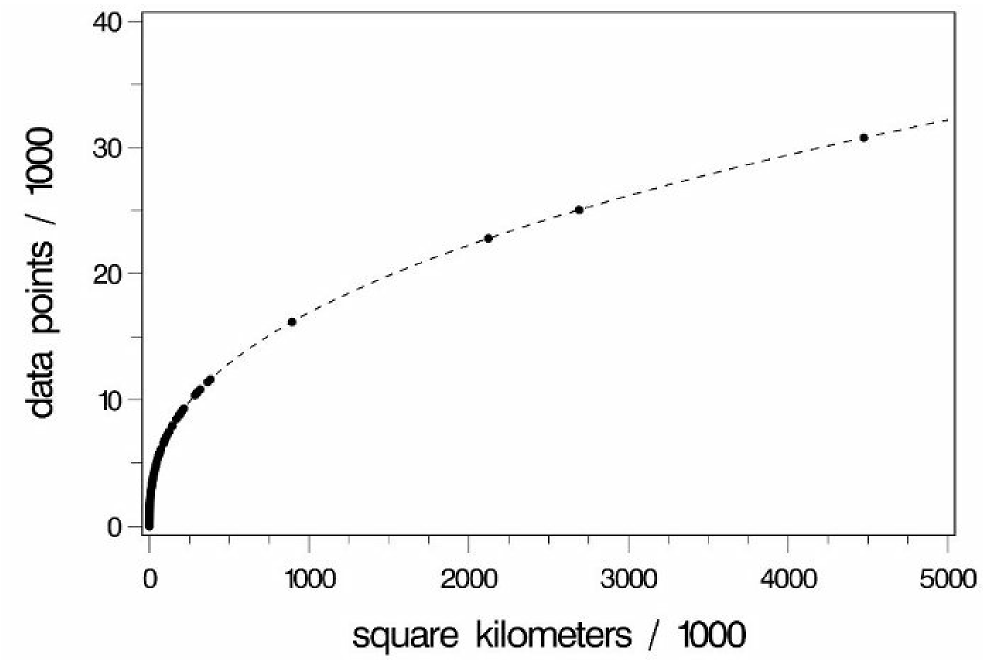

FIG. A1. Relationship between the number of data points extracted from a polygon in the digitized map of Brown et al. (1998) and the size of the polygon.

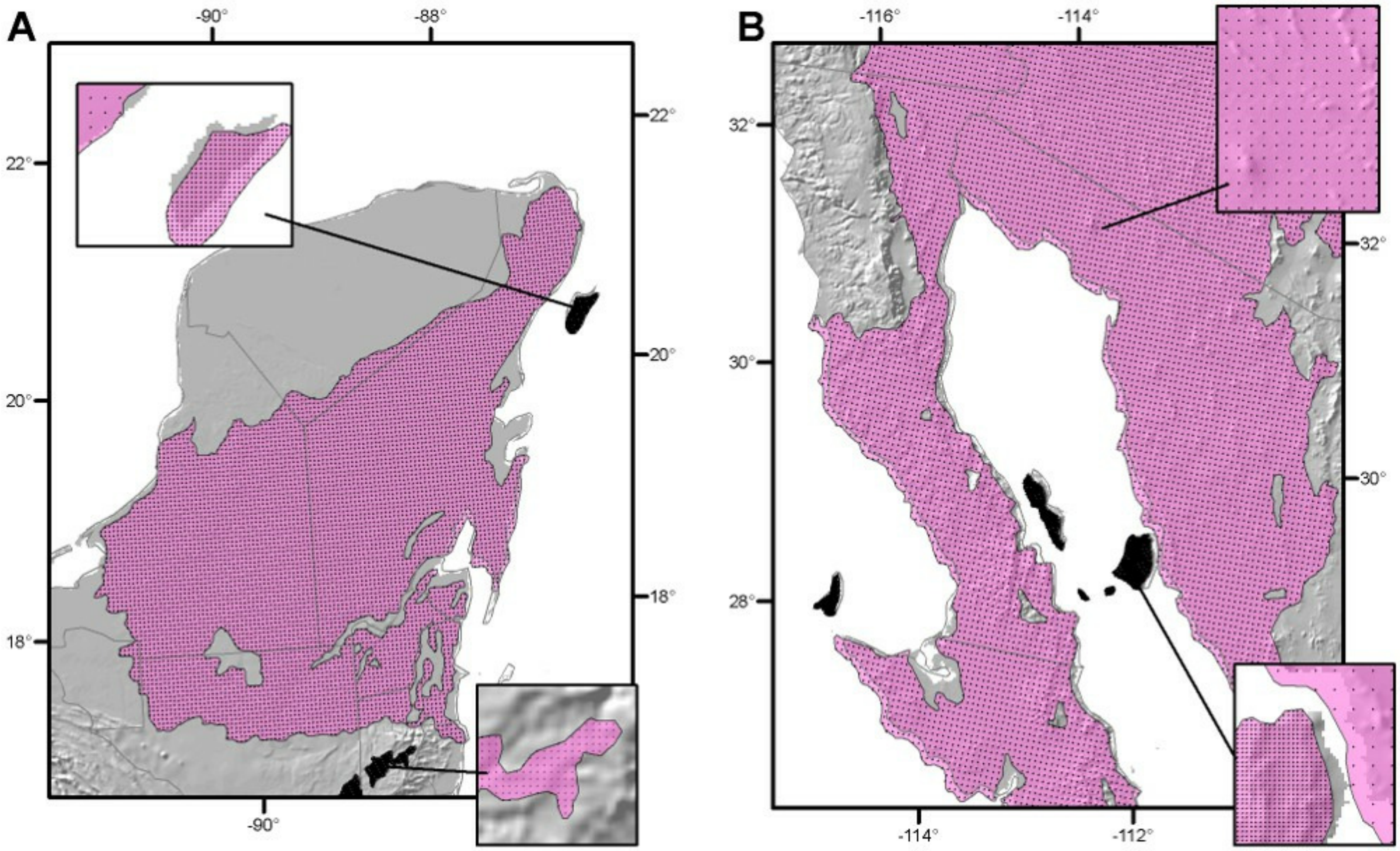

FIG. A2. Density of the sample (dots) for polygons of the Tropical Semi-Evergreen Forest of the Yucatan Peninsula of Mexico (A) and for the Sonoran Desertscrub (B) in southwestern USA and northern Mexico.

LITERATURE CITED

Arno, S. F. and R. P. Hammerly. 1984. Timberline. The Mountaineers. Seattle.

Brown, D. E, F. Reichenbacher, S. E. Franson. 1998. A classification of North American biotic communities. University of Utah Press, Salt Lake City. 141 pp.

Burns, R. M. and B. H. Honkala. 1990. Silvics of North America. 1. Conifers. Agriculture Handbook 654. Washington, DC: U.S. Department of Agriculture, Forest Service. 877p.

Daubenmire, R., and J. B. Daubenmire. 1968. Forest vegetation of eastern Washington and northern Idaho. Washington Agriculture Experiment Station Technical Bulletin 60. Washington State University, Pullman, Washington.

GLOBE Task Team. 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. http://www.ngdc.noaa.gov/mgg/topo/globe.html.

Hall, G. A. 1988. Birds of the southern Appalachian subalpine forests. Pages 101–118, In: J. A. Jackson ed. Bird Conservation 3. University of Wisconsin Press, Madison.

Ledig, F. T., G. E. Rehfeldt, C. Sáenz-Romero, C. Flores-López. 2010. Projections of suitable habitat for rare species under global warming scenarios. American Journal Botany 97:970–987.

Matos, J. A., 1995. *Pinus hartwegii* and *P. rudis*: A critical assessment. Systematic Botany 20, 6–21.

Meidinger, D. and J. Pojar. 1991. Ecosystems of British Columbia. Research Branch, British Columbia Ministry of Forests, Victoria, B.C., Canada.

R Development Core Team. 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rehfeldt, G. E. 2006. A spline climate model for western United States. General Technical Report 165. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Rehfeldt, G. E., N. L. Crookston, M. V. Warwell, and J. S. Evans. 2006. Empirical analyses of plant-climate relationships for the western United States. International Journal of Plant Science 167:1123–1150.

Sabo, S. R. and R. H. Whittaker. 1979. Bird niches in a subalpine forest: An indirect ordination. Proceedings National Academy Sciences 76:1338–1342.

Sáenz-Romero, C., and B. L. Tapia-Olivares, 2008. Genetic variation in frost damage and seed zone delineation within an altitudinal transect of *Pinus devoniana* (*P. michoacana*) in México. Silvae Genetica 57(3):165–170.

Sáenz-Romero, C, G. E. Rehfeldt, N. L. Crookston, P. Duval, and J. Beaulieu. 2010. Spline models of contemporary, 2030, 2060 and 2090 climates for México and their use in understanding climate-plant relationships. Climatic Change 102:595–623.

[Back to A022-007]