Supplementary Material S1: Detailed Description of the Derivation of Predictor Variables

Topo-climatic predictors

We applied a four-step procedure to create and then reduce the number of topo-climatic predictor variables considered. This process consisted of: (i) the generation of the topo-climatic predictor maps for zone 16; (ii) the extraction of the topo-climatic predictors by intersecting the FIA plot locations with each of the generated topo-climatic maps; (iii) a preliminary evaluation and subsequent transformation of the predictors for tree species modelling; and (iv) the reduction of the number of predictors to a small set of no more than 8, none of which had a Spearman rank correlation > 0.7. The four steps are described below.

We first generated a 90m digital elevation model (DEM) by re-sampling the 30m resolution national elevation data set (NED) (Gesch *et al.* 2002). Slope (SLP) was calculated in a GIS (ArcGIS, © Esri Inc. 2001), and topographic position (TOPO) was derived using the algorithm described in Zimmermann and Roberts (2001). This map represents a hierarchically integrated measured of standardized topographic exposure. The final map was smoothed by a 3x3 moving window.

Second, 90m bioclimatic parameter maps were generated from the DAYMET 1km daily gridded weather surfaces, available as 18-yr monthly and yearly climatological summaries[1]. DAYMET generates daily surfaces of temperature, precipitation, humidity, and radiation over large areas of complex terrain (Thornton *et al.* 1997). It uses a 1km DEM and observations of precipitation (PRCP), maximum and minimum temperature (TMAX, TMIN) from meteorological stations to generate maps of bioclimatic variables.

We down-scaled the 1km DAYMET monthly maps of TMIN, TMAX and PRCP to 90m resolution (see Zimmermann & Roberts 2001 for details) by first generating moving window regressions (radius ~20km) between the monthly 1km climate averages and the 1km DEM, and then writing the regression parameters to the centre cell of each 1km window position. The regression parameters (i.e., lapse rates and intercept) were next interpolated to a 90m resolution using inverse distance weighed interpolations. Finally, we generated monthly 90m maps for TMIN, TMAX and PRCP by applying the regression using the spatial maps of lapse rate, intercept and the 90m DEM. We next generated additional bioclimatic parameters based on the algorithms described in Thornton *et al.* (1997), including: (i) average daily and daytime temperature (TAVE, TDAY); (ii) ambient and saturated vapour pressure (VPAM, VPSA); and (iii) vapour pressure deficit (VPDD) and relative humidity (RELH). Additional monthly bioclimatic variables included: (i) degree-days of growing season (DDEG) using a threshold of 0°C; (ii) potential global radiation (SFMM) as the sum of potential direct and diffuse radiation based on the Kumar et al. (1997) method; (iii) potential evapo-transpiration (ETPJ) using the empirical equation of Jensen and Haise (1963); and (iv) a moisture index (MIND) as the difference between PRCP and ETPJ (see Zimmermann and Roberts 2001).

All FIA plots were next intersected with the 90m DAYMET-derived topo-climatic layers in a GIS. The selected topo-climatic variables were all hypothesized to have direct relationships to the distribution of the tree species to be modelled (Austin *et al.* 1984; Austin & Gaywood 1994). The intersected topo-climatic information (Table 1) was exported to the R statistical package (R Development Core Team 2004).

Preliminary analyses revealed that correlations among the monthly values for the 12 sets of bioclimatic predictors were high. Such extreme collinearity has implications for modelling. To minimize collinearity, a principal components analysis was carried out on each of the 12 sets of monthly bioclimatic predictors. In each case, the first principal component revealed the trend of the annual mean, while the second principal component showed a contrast of values between summer and winter months. For each set of 12 monthly variables, these two principal components explained over 88% of the variability, and in several cases the first two principal components explained over 99% of the variability in the sets of variables. Accordingly, for each set of monthly bioclimatic predictors we defined two new variables: (i) the yearly average of the 12 monthly variables; and (ii) the difference between the summer (July and August) and winter (December and January) months. In order to maximize the predictive power we defined summer as June-July for SFMM, and as August-September for RELH. Hereafter we use the suffix ‘‘*y*’’ to denote the average of the 12 monthly measurements, and the suffix ‘‘*d*’’ for the summer – winter difference. By this, we reduced the 12 x 12 (=144) monthly predictors to 24 possible predictor variables, plus SLP, TOPO and DDEG. These new predictors are easier to interpret than principal components, which carry similar information.

[1] http://www.daymet.org/

Last, we analysed the individual predictive power of the 27 remaining predictors on all 19 tree species. We were interested to further reduce the predictor set to a maximum of 8 predictors that enable accurate models for all tree species involved. Because we wanted to compare topo-climatic and remotely sensed predictors, both independent variable sets needed to be similar in the number of variables and include powerful predictors for all species involved. DDEG showed to be the best predictor across all species. Starting with DDEG, we added those predictors that had highest predictive powers on average for the 19 species if they correlated <0.7 (Green 1979) with the already selected ones. This was done to avoid the selection of highly correlated variables that do not add additional predictive power compared to the selected ones (Parra *et al.* 2004). The final eight selected topo-climatic predictors included: DDEG, TMIN.*d*, SFMM.*y*, RELH.*y*, PRCP.*y*, RELH.*d*, SLP, and TOPO.

Table 1: Description of topographic and monthly bioclimatic predictors analyzed for modelling the probability of presence for all tree species in the ZONE16 study area of the interior West. The lower case indices of the variable names used in the text (e.g. TMAX.*y*) indicate the aggregation to yearly (*y*) averages (sum for PRCP) and to differences (*d*) between the average of summer (July-August) and winter (December-January) months. DDEG was used as a yearly measure only. The column "selected" indicates what variable (*x*) or aggregation form (*y*, *d*) was selected for the topo-climatic model predictor set. The average correlation among the 8 selected model predictors is 0.20.

Type / Name	Description	selected	Units
Topographic			
SLP	slope	*x*	degrees,
0–90			
TOPO	topographic position	*x*	Index
Bioclimatic			
TMAX	monthly average maximum temperature	–	1/10 °C
TAVE	monthly average temperature	–	1/10 °C
TMIN	monthly average minimum temperature	*d*	1/10 °C
TDAY	monthly average daytime temperature	–	1/10 °C
DDEG	annual heat sum above 0°C	*x*	°C*days
PREC	precipitation sum	*y*	mm/mth
VPAM	ambient vapour pressure	–	Pa
VPSA	saturated vapour pressure	–	Pa
VPDD	vapour pressure deficit	–	Pa
RELH	relative humidity	*d*, *y*	%
SFMM	monthly potential global radiation	*y*	kJ/m2/day
ETPJ	potential evapotranspiration	–	mm
MIND	monthly moisture index	–	1/10 mm/day

Remotely sensed predictors

We used imagery obtained from the USGS Multi-Resolution Land Characteristics consortium of 2001[2] (MRLC 2001, hereafter) for our modelling. Imagery was collected for three different time periods representing the temporal dynamics of vegetation; early (spring), peak (summer), and late growing seasons (autumn). In order to distinguish between predictors of the respective seasons, we use the abbreviation "*.sp*", "*.su*", and "*.au*" hereafter. The MRLC processing standardization steps are summarized in the Landsat science data users handbook (Irish 2001), in Homer *et al.* (2004), and in the literature cited therein. Here, we briefly present the basic processing steps applied to the data set we used in our study area.

First all ETM+ bands were re-sampled using cubic convolution into an Albers Equal Area map projection. Then, the Tasselled Cap (TC) transformation (Kauth & Thomas 1976; Christ & Cicone 1984) – a linear re-combination of bands 1-5 & 7 – was applied according to Huang *et al.* (2002) resulting in three new products, namely the soil brightness index (SBI), the green vegetation index (GVI) and the wetness index (WI).

[2] http://www.mrlc.gov/

The normalized difference vegetation index (NDVI) was calculated using bands 3 (RED) and 4 (NIR), so that NDVI = (NIR+RED) / (NIR-RED). Finally, *L* of band 6 (thermal infrared) was converted to at-satellite temperature (T9 hereafter), providing a physically-based variable. Table 2 list all variables used and tested.

These indices plus bands 1-5 and 7 were then re-sampled to 90m based on 3x3 moving windows using the "average" and the "st.dev" function. This was done to (i) match the 90m spatial resolution of the topo-climatic predictors, and (ii) to cover an area that is at least the full spatial extent of the dependent forest inventory plot data. Next, all FIA plots used in this study were intersected with these layers in a GIS and the intersected remote sensing based predictors (Table 2) were exported to the R statistical package (R Development Core Team 2004).

Last, we examined the correlation structure and the individual predictive power of the 66 possible predictors (6 reflectance bands plus 5 transformed indices, both as 90m mean and standard deviation from images of three seasons each) on the 19 tree species used. These analyses revealed a high degree of correlation in the data set. We thus applied a similar pre-selection technique as for topo-climatic predictors for reducing the number of remote sensing based model predictors. The summer GVI was the best predictor overall. Starting with GVI.*su*, we added those predictors that had highest predictive powers overall, but had a Spearman rank correlation <0.7 with previously selected variables. We stopped the procedure after having selected 8 final remote sensing based model parameters, namely: GVI.*su*, WI.*sp*, .NDVI.*au*, T9.*su*, SBI.*su*, NDVI.*sp*, B3.*sp*, and T9.*au*. Only predictors originating from the "average" function in the 90m re-sampling remained in the model and spring band 3 was the only single band used for the modelling exercise. The final remote sensing based predictor set now had the same size as the topo-climatic predictor set.

Table S2: Description of the remotely sensed model predictors analyzed for modelling the probability of presence for all tree species in the ZONE16 study area of the interior West. Band 6 (thermal infrared at 10.4–12.5µm, 60m spatial resolution) was not used directly, but in the form of T9 (converted to at-sensor surface temperature). The 8 predictors selected in the model are listed with the season (*sp*=spring, *su*=summer, *au*=autumn), and with the resampling method (*avg*=average, *std*=standard deviation) used. The average correlation among the 8 selected model predictors is 0.39.

Type/Name	Description	selected method	season
Reflectance			
B1	Band 1; Visible Blue [0.450–0.515 µm]; 30m	–	–
B2	Band 2; Visible Green [0.525–0.605 µm]; 30m	–	–
B3	Band 3; Visible Red [0.630–0.690 µm]; 30m	*avg*	*sp*
B4	Band 4; Near Infrared [0.775–0.900 µm]; 30m	–	–
B5	Band 5; Shortwave Infrared [1.550–1.750 µm]; 30m	–	–
B7	Band 7; Shortwave Infrared [2.090–2.350 µm]; 30m	–	–
Recombinations			
SBI	Tasselled Cap Soil Brightness Index [using B1-B5, B7]	*avg*	*su*
GVI	Tasselled Cap Green Vegetation Index [using B1-B5, B7]	*avg*	*su*
WI	Tasselled Cap Wetness Index [using B1-B5, B7]	*avg*	*sp*
NDIV	Normalized Difference Vegetation Index [using B3, B4]	*avg*, *avg*	*sp*, *au*
T9	At Sensor Surface Temperature [using B6H]	*avg*, *avg*	*su*, *au*

References

Austin, M.P., Cunningham, R.B. & Fleming, P.M. (1984) New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, **55**, 11-27.

Austin, M.P. & Gaywood, M.J. (1994) Current problems of environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science*, **5**, 473-48.

Christ, E.P. & Cicone, R.C. (1984) A physically-based transformation of Thematic Mapper data – the TM Tasselled Cap, *IEEE Trans. on Geosciences and Remote Sensing*, **GE-22**, 256–263.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, & Tyler D. (2002) The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing*, **68**, 5–12.

Green, R.H. (1979). Sampling Design and Statistical Methods for Environmental Biologists. John Wiley, New York.

Homer, C., Huang, C.Q., Yang, L.M., Wylie, B. & Coan, M. (2004) Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, **70**(7), 829-40.

Huang, C., Wylie, B., Yang, L., Homer, C., & Zylstra, G. (2002) Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing*, **23**, 1741-1748.

Irish, R.R., 2001. Landsat 7 science data user's handbook, Report 430–15-01-003-0, National Aeronautics and Space Administration. [http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html]

Jensen, M.E. & Haise, H.R. (1963) Estimating evapotranspiration from solar radiation. *J. Irrig. Drainage Div. ASCE*, **89**, 15-41.

Kauth, R.J. & Thomas, G.S. (1976) The tasselled cap – a graphic description of the spectral-temporal development of agricultural crops as seen in Landsat, in *Proceedings on the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, Indiana, June 29 – July 1, 1976, (West Lafayette, Indiana: LARS, Purdue University), 41–51.*

Kumar, L., Skidmore, A.K., & Knowles, E. (1997). Modelling topographic variation in solar radiation in a GIS environment. *International Journal for Geographical Information Science,* **11**, 475-497.

Parra, J.L., Graham, C.C. & Freile, J.F. (2004) Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, **27**(3), 350-60.

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

Thornton, P.E., Running, S.W. & White, M.A. (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**(3-4), 214-51.

Zimmermann, N.E. & Roberts, D.W. (2001) Final report of the MLP climate and biophysical mapping project. Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. [http://www.wsl.ch/staff/niklaus.zimmermann/mlp/mlp_report.pdf]