ORIGINAL ARTICLE

WILEY **Journal of Biogeography**

Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species

Fernando O. G. Figueiredo[1,2] (iD) | Gabriela Zuquim[3] | Hanna Tuomisto[3] | Gabriel M. Moulatlet[3] | Henrik Balslev[4] | Flávia R. C. Costa[2]

[1]Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia—INPA, Manaus, Brazil

[2]Coordenação de Pesquisas em Biodiversidade, Instituto Nacional de Pesquisas da Amazônia—INPA, Manaus, Brazil

[3]Department of Biology, University of Turku, Turku, Finland

[4]Department of Bioscience, Aarhus University, Aarhus C, Denmark

Correspondence
Fernando O. G. Figueiredo, Instituto Nacional de Pesquisas da Amazônia Av. André Araújo, 2.936 - Petrópolis - CEP 69.067-375 - Manaus - Amazonas Brasil
Email: nandoeco06@gmail.com

Editor: Steven Higgins

Abstract

Aim: To evaluate the relative importance of climatic versus soil data when predicting species distributions for Amazonian plants and to gain understanding of potential range shifts under climate change.

Location: Amazon rain forest.

Methods: We produced species distribution models (SDM) at 5-km spatial resolution for 42 plant species (trees, palms, lianas, monocot herbs and ferns) using species occurrence data from herbarium records and plot-based inventories. We modelled species distribution with Bayesian logistic regression using either climate data only, soil data only or climate and soil data together to estimate their relative predictive powers. For areas defined as unsuitable to species occurrence, we mapped the difference between the suitability predictions obtained with climate-only versus soil-only models to identify regions where climate and soil might restrict species ranges independently or jointly.

Results: For 40 out of the 42 species, the best models included both climate and soil predictors. The models including only soil predictors performed better than the models including only climate predictors, but we still detected a drought-sensitive response for most of the species. Edaphic conditions were predicted to restrict species occurrence in the centre, the north-west and in the north-east of Amazonia, while the climatic conditions were identified as the restricting factor in the eastern Amazonia, at the border of Roraima and Venezuela and in the Andean foothills.

Main conclusions: Our results revealed that soil data are a more important predictor than climate of plant species range in Amazonia. The strong control of species ranges by edaphic features might reduce species' abilities to track suitable climate conditions under a drought-increase scenario. Future challenges are to improve the quality of soil data and couple them with process-based models to better predict species range dynamics under climate change.

KEYWORDS
Amazon rain forest, Bayesian logistic regression, cation exchange capacity, climate change, ecological niche models, soil factors, SoilGrids, species distribution models, species range, tropical soils

1 | INTRODUCTION

Climate is an important determinant of species ranges at broad spatial scales, so an understanding of how climate change will affect ecological systems is urgently needed. The most obvious effect that climate change can be expected to have on biodiversity is to shift in species' geographical distributions (Parmesan & Yohe, 2003; Thuiller, 2007). To avoid extinction, species must either track suitable climate or adapt to novel climatic conditions. Some species are favoured by climate changes and expand their ranges, but range shifts may be decoupled from climatic tracking by other factors that restrict species distributions, such as non-environmental spatial constraints, land use, biotic interactions and soil conditions (Blach-Overgaard, Svenning, Dransfield, Greve, & Balslev, 2010; Eiserhardt, Svenning, Kissling, & Balslev, 2011; Hayes & Sewlal, 2004; Marage & Gégout, 2009; Wisz et al., 2013). Identification of both climatic and non-climatic constraints on broad-scale species distributions is therefore necessary to better predict how species will respond to climate change.

Recent projections of global climate models predict significant climate change in Amazonia by the end of the 21st century (Boisier, Ciais, Ducharne, & Guimberteau, 2015), which would affect the plant diversity patterns in important ways (Olivares, Svenning, van Bodegom, & Balslev, 2015). It is already known that soil conditions affect species occurrence patterns in Amazonia, and therefore they deserve special attention in the modelling of plant distributions. Edaphic heterogeneity is tightly related to geological features, such as soil age and mineralogical composition of the parent material (Irion, 1978; Quesada et al., 2010). Although climate is relatively uniform over large parts of the Amazon basin, geological formations differ widely in age and geochemistry. Many studies suggest that soils exert the main control on broad-scale floristic patterns in Amazonia (Fittkau, Junk, Klinge, & Sioli, 1975; Higgins et al., 2011; Sombroek, 2000; Tuomisto & Poulsen, 1996). However, studies focused on the southern and northern extremes of Amazonia show significant turnover of species along precipitation and temperature gradients (ter Steege et al., 2006; Toledo et al., 2011) suggesting that climatic conditions may be important in constraining species ranges only at the borders of the biome. Although both climate and soil are considered important determinants of broad-scale floristic patterns in Amazonia, there has been no evaluation of their relative contributions to predicting plant species ranges.

Several studies have used spatial correlative models to predict current species distributions (Buermann et al., 2008; Moscoso, Albernaz, & Salomão, 2013; Vedel-Sørensen, Tovaranonte, Bøcher, Balslev, & Barfod, 2013) and potential range shifts under climate change for Amazonian species (Miles, Grainger, & Phillips, 2004; Feeley, Malhi, Zelazowski, & Silman, 2012; Thomas, Caicefo, Loo, & Roeland, 2014; Thomas et al., 2012). However, either these studies relied on climatic and remote sensing variables, or used proxies, such as coarse-resolution polygon-based maps of ecoregions or soil types, to represent environmental factors other than climate (Feeley et al., 2012; Thomas et al., 2014; Vedel-Sørensen et al., 2013). Soil variables are often neglected in species distribution models (SDM) because accurate and ecologically relevant data covering large extents are scarce. Recent efforts to develop digital soil maps, such as SoilGrids, have increased the availability of high-resolution grid-based soil information (Hengl et al., 2014), which include chemical and physical properties of the soil, as well as occurrence probabilities of soil classes. These new soil products may predict plant species range better than the previously used coarse-resolution proxies.

Here, we present the first broad-scale evaluation of how soil and climate variables interplay to predict plant species ranges in Amazonia using SDMs and a diverse set of climate and soil data layers. We focused on 42 plant species from phylogenetically distant groups that represent different growth forms: trees, palms, lianas, ferns and monocot herbs. We aimed to: (1) test whether the inclusion of soil variables improves the performance of climate-based SDMs; (2) quantify the unique and joint contributions of climate and soil variables in explaining species distributions; (3) quantify the relative importance of individual variables in predicting species distributions; and (4) predict areas where soil and climatic conditions, independently and jointly, restrict species ranges. We expect that inclusion of a physiologically relevant set of edaphic predictors in SDM in combination with climate variables would produce more accurate predictions of species distributions and shed light on how soil and climate changes may interact to drive shifts in species ranges in Amazonia.

2 | MATERIALS AND METHODS

2.1 | Species and Environmental data

2.1.1 | Species data

We focused on 42 plant species (Table 1) to investigate the predictive performance of climate and soil variables for mapping species distributions. The species were selected to cover a wide spectrum of plant sizes and life history strategies (trees, palms, lianas, monocot herbs and ferns) as well as different range sizes (some restricted to Amazonian rain forests and others also found in other biomes). Species occurrence records were obtained from two sources: plot-based inventories and herbarium records. To ensure data consistency, we targeted species that are easy to identify in the field. We included only species that had more than 20 presence records (see further details: Table S1, Appendix S1 in Supporting Information).

2.1.2 | Environmental data

As climate predictors, we used both WorldClim variables and remote sensing data. Out of the available WorldClim data (http://www.worldclim.org), we used the 19 bioclimatic variables, which express 11 temperature and 8 precipitation metrics at about 1-km resolution (WorldClim version 1.4; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). The WorldClim variables were produced by spatial interpolation between meteorological stations, which creates high

TABLE 1 List of the 42 Amazonian plant species, geographical regions of occurrence and the number of occurrence records (N) after removing outliers and applying a spatial filtering procedure with the indicated radius distance (d) to reduce sample bias. The geographical regions were classified as Amazonian rain forest (AM), Central American rain forest (CA), Cerrado biome (CE) and Atlantic rain forest (AT)

Plant group	Species	Family	Geographical distribution	N	d (km)
Tree	*Caryocar glabrum* (Aubl.) Pers.	Caryocaraceae	AM	375	50
Tree	*Caryocar microcarpum* Ducke	Caryocaraceae	AM	106	25
Tree	*Couepia dolichopoda* Prance	Chrysobalanaceae	AM	24	5
Tree	*Eperua falcata* Aubl.	Fabaceae	AM	72	50
Tree	*Henriquezia nitida* Spruce ex Benth.	Rubiaceae	AM	43	5
Tree	*Inga alba* (Sw.) Willd.	Fabaceae	AM, CA, CE, AT	534	50
Tree	*Jacaranda copaia* (Aubl.) D. Don	Bignoniaceae	AM, CA	513	50
Tree	*Jacaratia spinosa* (Aubl.) A. DC.	Caricaceae	AM, CA, CE, AT	238	50
Tree	*Manilkara huberi* (Ducke) A. Chev.	Sapotaceae	AM	178	50
Tree	*Mezilaurus itauba* (Meisn.) Taub. ex Mez	Lauraceae	AM	318	50
Tree	*Minquartia guianensis* Aubl.	Olacaceae	AM, CE	183	50
Tree	*Nectandra turbacensis* (Kunth) Nees	Lauraceae	AM, CA, CE	89	50
Tree	*Poecilanthe effusa* Huber	Fabaceae	AM	154	50
Tree	*Simarouba amara* Aubl.	Simaroubaceae	AM, CA, CE, AT	576	50
Tree	*Siparuna decipiens* (Tul.) A. DC.	Siparunaceae	AM	411	50
Palm	*Astrocaryum aculeatum* G. Mey.	Arecaceae	AM	96	50
Palm	*Bactris acanthocarpa* Mart.	Arecaceae	AM, AF	108	50
Palm	*Iriartea deltoidea* Ruiz & Pav.	Arecaceae	AM.CA	133	50
Palm	*Leopoldinia pulchra* Mart.	Arecaceae	AM	45	15
Palm	*Oenocarpus bataua* Mart.	Arecaceae	AM	106	50
Palm	*Socratea exorrhiza* (Mart.) H. Wendl.	Arecaceae	AM, CA	210	50
Liana	*Callichlamys latifolia* (Rich.) K. Schum.	Bignoniaceae	AM, CA, CE, AT	231	50
Liana	*Machaerium amplum* Benth.	Fabaceae	AM, CE, AF	83	20
Liana	*Machaerium ferox* (Mart. ex Benth.) Ducke	Fabaceae	AM	62	5
Liana	*Machaerium multifoliolatum* Ducke	Fabaceae	AM	41	15
Liana	*Martinella obovata* (Kunth) Bureau & K. Schum.	Bignoniaceae	AM, CA, CE	162	50
Liana	*Pleonotoma jasminifolia* (Kunth) Miers	Bignoniaceae	AM, CE	83	15
Liana	*Rourea cuspidata* Benth. ex Baker	Connaraceae	AM, CE	74	30
Monocot herb	*Goeppertia fragilis* (Gleason) Borchs. & S. Suárez	Marantaceae	AM	20	5
Monocot herb	*Goeppertia loeseneri* (J.F. Macbr.) Borchs. & S. Suárez	Marantaceae	AM	33	5
Monocot herb	*Goeppertia zingiberina* (Körn.) Borchs. & S. Suárez	Marantaceae	AM	69	20
Monocot herb	*Heliconia schumanniana* Loes.	Heliconiaceae	AM	93	10
Monocot herb	*Hylaeanthe hexantha* (Poepp. & Endl.) A.M.E. Jonker & Jonker	Marantaceae	AM	45	5
Monocot herb	*Ischnosiphon martianus* Eichler ex Petersen	Marantaceae	AM	28	30
Monocot herb	*Monotagma ulei* K. Schum.	Marantaceae	AM	30	5
Monocot herb	*Renealmia breviscapa* Poepp. & Endl.	Zingiberaceae	AM	80	50
Fern	*Adiantum pulverulentum* L.	Pteridaceae	AM, CA, CE, AT	257	50
Fern	*Adiantum tomentosum* Klotzsch	Pteridaceae	AM	120	40
Fern	*Cyathea pungens* (Willd.) Domin	Cyatheaceae	AM, CA, CE, AT	168	50
Fern	*Cyclopeltis semicordata* (Sw.) J. Sm.	Lomariopsidaceae	AM, CA	190	50
Fern	*Lindsaea guianensis* (Aubl.) Dryand.	Lindsaeaceae	AM, CA, CE, AT	100	50
Fern	*Schizaea elegans* (Vahl) Sw.	Schizaeaceae	AM, CA, CE, AT	199	50

uncertainty in the estimated climate data in many parts of Amazonia, where stations are sparse. Therefore, we also included precipitation data estimated by the Tropical Rainfall Measuring Mission (TRMM)

satellite (Kummerow, Barnes, Kozu, Shiue, & Simpson, 1998; http://disc.sci.gsfc.nasa.gov). We converted monthly data from 1998 to 2004 of the TRMM product 3B43 V6 at a 0.25° resolution (about

28 km at the equator) into two climatic variables: annual precipitation and dry season length, defined as the maximum consecutive number of months with <100 mm of precipitation.

We used four remote sensing variables that describe terrain and forest structure properties: elevation, percentage tree cover, percentage evergreen tree cover and canopy height. Elevation was obtained from the Shuttle Radar Topography Mission (http://worldgrids.org). Tree cover and evergreen tree cover were derived from images of the NOAA's Advanced Very High Resolution Radiometer acquired in 1992–1993 (DeFries, Hansen, Townshend, Janetos, & Loveland, 2000; http://www.landcover.org). Canopy height was derived from the spaceborne Geoscience Laser Altimeter System (Simard, Pinto, Fisher, & Baccini, 2011; http://daac.ornl.gov). All remote sensing variables were originally provided at ~1-km resolution. We included the terrain and forest structure data in the pool of climate variables because these variables can be expected to represent ecophysiological constraints on species distribution via complex interactions between temperature, precipitation and solar radiation (Grubb, 1977; Kempes, West, Crowell, & Girvan, 2011).

Soil variables were obtained from the SoilGrids 250 m database (https://www.soilgrids.org). We used 19 soil variables that provide predicted values for the surface soil layer (0–15 cm depth). These included: three soil texture variables (percentages of clay, sand and silt), which are related to soil drainage; two chemical variables related to nutrient availability, the cation exchange capacity (CEC, measured in $cmol_c$ kg^{-1} at pH 7) and pH (determined in KCl); and other 14 variables representing different soil classes (following the classification of FAO's World Reference Base), which are related to soil physical conditions and nutrient availability (Quesada et al., 2010). The soil classes were selected based on their moderate to high occurrence probabilities in the Neotropics. We used each soil class layer as an independent continuous predictor, where the value for a given grid cell represents the probability of occurrence of that soil class. In this way, our SDMs account for the uncertainty associated with the spatial limits of soil classes, and contrasts markedly with approaches that treat polygon-based soil classes as categorical variables in statistical models, where each grid cell belongs to one soil class only. Extreme environmental conditions (bare soils, sparsely vegetated areas and water bodies), which are unlikely to be occupied by our species were excluded by combining the spatial mask present in all environmental data and the spatial mask employed in SoilGrids 1 km (Hengl et al., 2014; product SMKISR accessed in http://worldgrids.org). The latter excludes several non-vegetated areas that are included in SoilGrids 250 m. We rescaled all environmental data to a 0.05° spatial resolution. Both environmental and species data were transformed to the Mollweide equal-area projection before analysis. The maps of all 44 environmental variables are presented in Figure S1.1.

2.2 | Modelling framework

To evaluate the role of climate and soil factors in predicting species ranges, we built a set of three SDMs for each species: (1) CLIM models based on WorldClim and remote sensing variables; (2) SOIL models based on SoilGrids variables; (3) and CLIM+SOIL models based on both sets of variables together. Our modelling approach includes six steps, as detailed below:

Step 1: Pre-modelling: detecting outliers, reducing sample bias and setting geographical background.

To improve SDM quality, we first coarsened the spatial resolution of the occurrence data. We retained only one presence record per 5 × 5 km grid cell and removed geographical and environmental outliers in the presence records. Then, we applied a spatial filter procedure to reduce sample bias (spatial aggregation) of occurrence records (Figure S1.2). Finally, we constrained the area for model calibration, evaluation and prediction separately for each species to avoid predicting species presence in areas far beyond their potential ranges. Details of the pre-modelling procedures are given in Appendix S1b.

Step 2: Predicting relative species occurrence probabilities using Bayesian logistic regression

To predict the relative occurrence probability of a species in each 5 × 5 km grid cell, we applied logistic regression using presence and background points. Logistic regression using presence and background points is similar to methods widely used in ecology (e.g. MAXENT and resource selection functions). All of these methods are closely related to the Poisson point process model recently viewed as an appropriate way to analyse presence-only data in a regression framework (Aarts, Fieberg, & Matthiopoulos, 2012; Renner & Warton, 2013; Renner et al., 2015; Warton & Shepherd, 2010). For each species, we generated 10,000 random background points over the geographical background area defined in step 1. We used Bayesian logistic regression instead of a conventional maximum likelihood logistic model because the maximum likelihood algorithm tends to overestimate β coefficients when the number of presence records is very small in relation to absence or background points, or when the distribution of a species along an environmental gradient is very narrow (Hefley & Hooten, 2015). Bayesian logistic regression circumvents these problems by combining the information of the likelihood functions and a weakly informative default prior based on the Cauchy distribution to estimate model parameters (Gelman, Jakulin, Pittau, & Su, 2008).

We performed stepwise model selection (backward–forward) using the Akaike information criterion (AIC) to find the most parsimonious model for each species separately. We allowed high correlated variables belonging to different source groups (WorldClim, TRMM climatic data, remote sensing data or SoilGrids) to compete in the same model selection. Before the model selection procedure, we found pairs of variables that had correlation greater than 0.7 (absolute value of Pearson correlation) and excluded the one with greater variance inflation factor. We stopped the process when all pairwise correlations were less than 0.7. Only linear and second-order polynomial effects (Gaussian-shaped curve) were used. We started the backward–forward procedure including linear effects only. Second-order polynomials were allowed to enter the models at any step. We repeated model selection three times for each species to obtain the best model of CLIM+SOIL, CLIM and SOIL model.

Step 3: Evaluating model performance

To evaluate the performance of CLIM+SOIL, CLIM and SOIL models, we employed the ΔAIC, i.e. the difference in AIC between the model being evaluated and the best model (Burnham & Anderson, 2002). As a rule of thumb, ΔAIC values smaller than 2 indicate models with good support. Null models—those with only the intercept term—were also included in the evaluation.

We evaluated the spatial accuracy of models using the area under the receiver operating characteristic curve (AUC; Bradley, 1997). For the best models returned by the model selection procedure, we applied a 10-fold cross-validation and computed the AUC test of each replicate. AUC values range from 0 to 1, with AUC values higher than 0.7 indicating models with good performance and values lower than 0.5 indicating that a model is worse than the null model (Araújo, Pearson, Thuiller, & Erhard, 2005). Before computing AUC, the relative occurrence probability (ROP) was transformed into cumulative values (hereafter defined as suitability index), which range from 0 to 100 and is better than the raw ROP output to visualize species range boundaries (Merow, Smith, & Silander, 2013; Phillips, Anderson, & Schapire, 2006). Final prediction of all models was produced by averaging the 10 replicates returned by cross-validation. We applied a two-tailed nonparametric Wilcoxon test to check if spatial prediction's accuracy measured by the AUC was higher than that of null models. For those species with best models better than null, the difference in AUC scores between CLIM+SOIL, CLIM and SOIL models was tested using generalized mixed effect models (GLMM) (Pinheiro & Bates, 2000) with species as the random term and assuming that AUC values follow a beta distribution.

Step 4: Quantifying the unique and joint effects of climate and soil variables on species distribution models

As Bayesian logistic regression can be viewed as a generalized linear model with a constrained approach to estimate model parameters, we employed the same procedure of variance partitioning used in linear modelling (Borcard, Legendre, & Drapeau, 1992) to quantify how much of the variances explained by the best CLIM+SOIL, CLIM and SOIL models were explained uniquely by climate or soil variables and jointly by both sets of variables. Given that we applied the model selection procedure to find the best models in each of the three model settings, the SOIL and CLIM models do not necessarily include the same subset of variables that were retained in the best CLIM+SOIL model. Hence, our approach is not a true variance partitioning, but an approximation of it. Percentage of total explained variance (EV) in logistic models, also known as pseudo-R^2, was given by:

$$EV = \frac{\text{null deviance} - \text{residual deviance}}{\text{null deviance}} \times 100$$

as suggested by Zuur, Ieno, Walker, Saveliev, & Smith (2009). The unique contribution of climate was computed as $EV_{clim+soil} - EV_{soil}$, and that of soil as $EV_{clim+soil} - EV_{clim}$. The joint contribution of climate and soil was obtained as $EV_{soil} + EV_{clim} - EV_{clim+soil}$. We report the unique and joint contributions of CLIM and SOIL model as percentages of EV rather than as percentages of total variance. This is

done because background data may be generated also for pixels containing presence records, so the total variance will never be fully explained.

Step 5: Quantifying the relative importance of individual variables.

For each species, we ran Bayesian logistic regressions separately with each variable to quantify their relative importance. For each of 44 predictors, we ran single-variable SDM and selected the best shape of the species response curve (no effect, linear or second-order polynomial) using ΔAIC scores. The EV of the single-predictor model was then expressed as a percentage of the EV of the best model identified in step 3. Finally, we ranked variables based on their median EV values over all species.

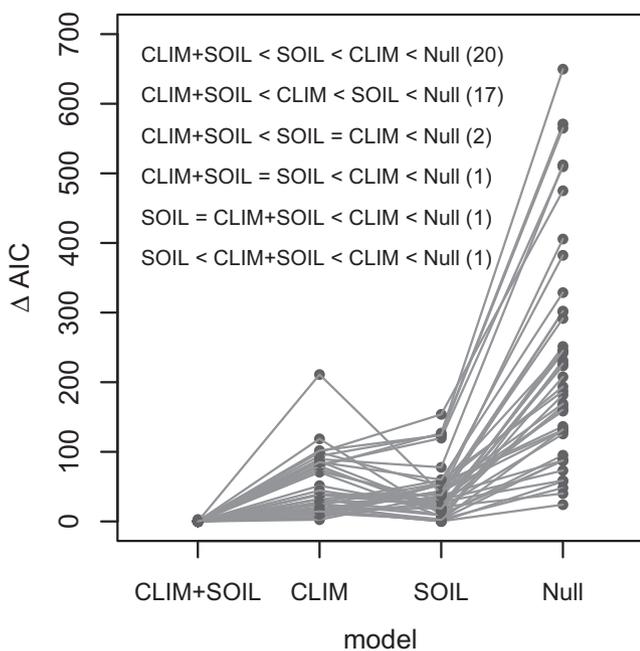*Step 6: Mapping the difference between predictions of CLIM and SOIL models.*

We mapped the difference between the models predictions only in areas defined as unsuitable, i.e. in areas that species are likely to be absent. To define the unsuitable areas, we applied a threshold on each species suitability scores that maximizes the sum of sensitivity (1—omission error) and specificity (the fraction of predicted area) (Manel, Williams, & Ormerod, 2001). First, we mapped the difference in suitability between the best model and each of the models being tested, the CLIM or SOIL models. In both case, high values (over-predictions) indicate areas where the variables evaluated are unlikely to be the restricting factor for species occurrence. Second, to identify areas where both predictors may jointly or independently restrict species range, we mapped the difference between predictions of the CLIM and SOIL model. Values close to zero indicate areas where we hypothesized that both soil and climate restrict species range. Positive values indicate regions where the CLIM model over-predicted suitable conditions relative to SOIL model and, hence, we hypothesized that species range was mainly restricted by soil in such areas. Conversely, negative values indicate regions where climate may be the main restricting factor. Finally, we averaged the maps of all species across Amazonia sensu latissimo (Eva & Huber, 2005) to produce an overall predictive map of potential climate and soil constraint on species ranges.

All analyses were conducted using the R platform (3.3.2; R Development Core Team, 2015). Manipulation of raster and vector data was done using the 'raster' (Hijmans, 2016) and 'sp' (Pebesma & Bvand, 2005) packages. Data from GBIF and iDigBio were downloaded using the function *occ* from the package 'spocc' (Chamberlain, Ram, & Hart, 2016). Geographical outliers were detected and removed using the function *rjack* of the 'biogeo' package (Robertson, 2016). We executed the spatial filtering procedure using functions *lohboot* and *remove.duplicates* of the 'spatstat' (Baddeley & Turner, 2005) and 'dismo' packages (Hijmans, Phillips, Leathwick, & Elith, 2017), respectively; exclusion of correlated variables with *vifcor* function of the 'usdm' package (Naimi, 2015); Bayesian logistic regression with function *bayesglm* of the 'arm' package with default options (Gelman & Su, 2016); computation of AUC scores using function *auc* from package 'MESS' (Ekstrøm, 2016); and evaluation of model significance with the *wilcox.test* function of the 'stats' package (R Development Core Team, 2015). The tests of differences in model

performance were done fitting a GLMM using function *gamlss* of the 'gamlss' package (Rigby & Stasinopoulos, 2005). We implemented our own R function to perform the model selection procedure, which is available in Appendix S1c.
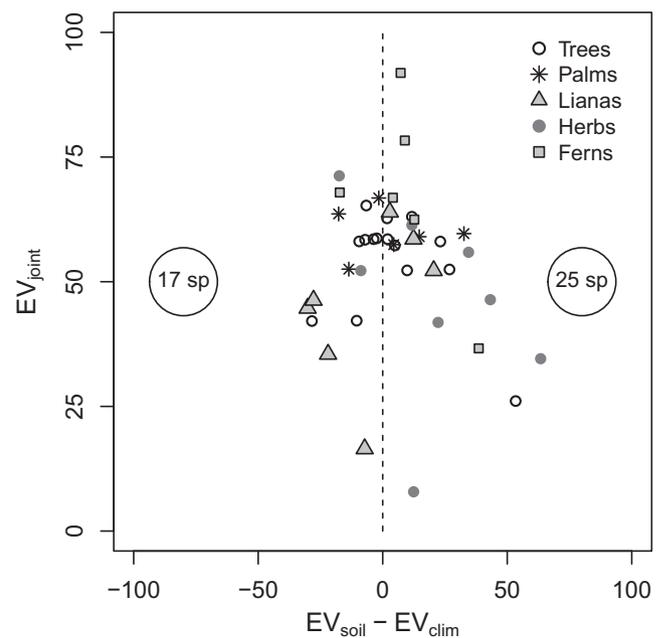
3 | RESULTS

Our results show that soil data are important for the species distribution models developed in this study (Figure 1). For 39 of the 42 species, the models combining climatic and soil data (CLIM+SOIL) were supported as the best models. In the other three species, either the relative ranking of the CLIM+SOIL and SOIL models was uncertain, or the SOIL model had more support. The CLIM model was not supported as the best model for any of the species. Models including soil predictors only had better performance than models including climate predictors only. Across all species the AUC values for the SOIL models were significantly lower than those for CLIM+SOIL models ($p < .001$) but higher than those for CLIM models ($p < .005$). For every species, the best models showed fair to good prediction accuracy. All median AUC of the best models were greater than 0.7 and significantly greater than the null model (all $p < .005$ based on Wilcoxon test of AUC). All model statistics (ΔAIC, AUC, omission rates, fractions of predicted area and thresholds used to define unsuitable areas) are provided in Figure S1.3.

For 25 species (60%), more variance was explained uniquely by SOIL variables than uniquely by CLIM variables, and for 30 species (71%), more than 50% of the explained variance was due to the joint contribution of CLIM and SOIL variables (Figure 2). Annual precipitation, dry season length, canopy height, tree cover and Haplic Xanthic Ferralsols were the variables with highest contributions to explained variance, on average, in the single-variable SDMs (Figure 3). However, single-variable models never explained more than 15% of the variance explained by the corresponding best model, and the average was <8% for each one of the 44 variables. Species response curves along the most important CLIM variables revealed that most species tend to avoid dry conditions (drought-sensitive) and occur preferentially in tall closed forest (Figure S1.4).

Visual assessment of the SDMs for each of the 42 species showed that CLIM+SOIL models predicted more patchy suitable areas than CLIM models (Figure S1.3). CLIM models over-predicted suitable conditions in central Amazonia, upper Rio Negro, northern Guianas and in some areas of southern Amazonia, which reveals that it is more likely that species are absent in these areas due to soil restriction, not due to climate restriction (Figure 4a). The opposite was true in areas where the SOIL model over-predicted suitable conditions, i.e. in the Andean foothills, the border zone between Peru and Acre state, eastern Amazonia, and the savannas of Roraima and Venezuela (Figure 4b). Both models predicted low suitability in


FIGURE 1 Performance of CLIM+SOIL, CLIM and SOIL species distribution models of 42 plant species in Amazonia. The ranking of model performance is based on ΔAIC scores, and the number of species with each kind of ranking is shown in parentheses. The "=" indicates a difference smaller than 2 (no evidence that the model with lower ΔAIC is better) and "<" indicates a difference greater than 2 (evidence that model with lower ΔAIC is better). ΔAIC values for each species separately are presented in Figure S1.3

FIGURE 2 Contribution to explained variance (EV) of the unique and joint fractions of the soil and climate models for each of 42 Amazonian plant species (species of different life-forms are shown with different symbols). EV_{joint} represents the EV that is jointly explained by soil and climate variables. Horizontal axis represents the difference between the unique contributions of the soil model (EV_{soil}) and climate model (EV_{clim}). Species for which the unique contribution of the soil model was higher than the unique contribution of the climate model are found to the right of the vertical midline
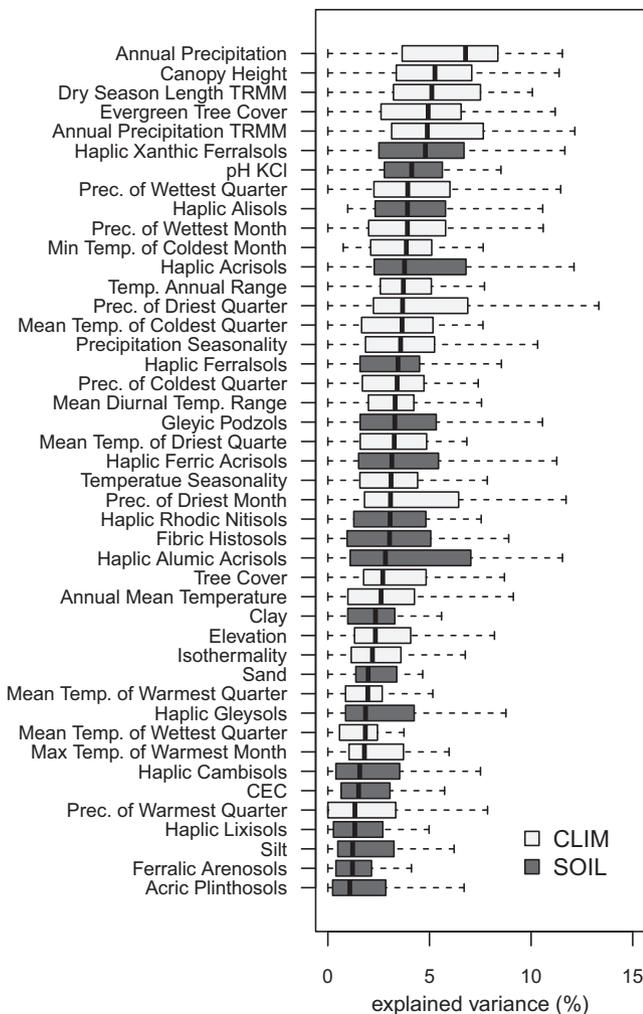
FIGURE 3 Individual variables ranked by their importance in SDMs of 42 plant species in Amazonia. Importance is defined as the median value across all species of the explained variance (EV) in single-variable models, expressed as a percentage of the EV in the corresponding best model. Variables are coloured according to whether they were used in the CLIM models or in the SOIL models

southern Amazonia, which indicates that soil and climate jointly restrict species occurrence in this area (Figure 4c).

4 | DISCUSSION

Using SDM and a diverse set of climatic and edaphic predictors, we demonstrated that the soil-only models provide more accurate predictions than the climate-only models. This can be interpreted as contradicting the classical assumption of biogeography that climate is the main driver of broad-scale species distribution patterns (Gaston, 2003; Pearson & Dawson, 2003). The higher importance of soil relative to climate may reflect the fact that the variation in rainfall and temperature is narrow across large parts of Amazonia. On the other hand, the edaphic differences among regions can be so large that they act as filters preventing species from occurring over the

entire biome. Indeed, the combined soil and climate models had the best predictive power for most species, which is consistent with earlier findings suggesting that both climate and soil properties are important as drivers of plant diversity and community composition across Amazonia (ter Steege et al., 2006; Tuomisto, Ruokolainen, & Yli-Halla, 2003; Zuquim et al., 2014).

Climate and soils have been invoked to explain spatial patterns of floristic variation across Amazonia. The main axis of tree floristic variation has been associated with the south-west/north-east soil fertility gradient and the second axis with a north-west/south-east precipitation gradient (ter Steege et al., 2006). Our predictive maps reveal a more complex and slightly different spatial pattern of soil and climate control on species distribution. Soil might be more decisive than climate in restricting species occurrence particularly in the upper Rio Negro area, central Amazonia and the northern part of the Guiana region, which are covered predominantly by poor nutrient soils (Quesada et al., 2010; Sombroek, 2000; Zuquim et al., 2014). These findings are in accordance with previous studies that predict the absence of species associated to rich soils in large areas of Amazonia, particularly in the centre of biome (Tuomisto et al., 2016; Zuquim et al., 2014). On the other hand, where climate conditions vary markedly over short distances (in the Andean foothills, in the border between Peru and Acre state, eastern Amazonia and the savannas of Roraima and Venezuela), climate may become more restrictive to species occurrence, especially if soil conditions are relatively uniform. The southern and north-western borders of Amazonia were predicted as unsuitable by both climate-only and soil-only models, suggesting that the north-west/south-east floristic gradient interpreted as climatic by ter Steege et al. (2006) may also reflect soil-related factors.

When the environmental variables were used separately in single-predictor SDMs, the 10 most important predictors included precipitation-related variables, forest structural variables and soil variables. However, alone they explained <15% of the variance explained by the corresponding best model, indicating that species ranges are controlled by multiple factors. Indeed, for most species, the proportion of deviance explained jointly by climatic and edaphic variables was more than 50% of total explained variance, suggesting some degree of correlation between our predictor variables. Such relationships can be causal, as temperature, precipitation and topography are among the main factors that affect soil formation (Jenny, 1994). Nevertheless, climate conditions seem to be less important than age and quality of the parent material and dynamic process of removal-deposition of sediments in determining soil properties in Amazonia (Irion, 1978; Quesada et al., 2010; Räsänen, Salo, & Kalliola, 1987). Forest structure and elevation, used here as climate variables, may also be inter-correlated with soil properties, such as soil texture and nutrient content (Costa, Magnusson, & Luizão, 2005; Higgins et al., 2011; Levine et al., 2016; Tuomisto, Ruokolainen, Aguilar, & Sarmiento, 2003). More detailed studies are still needed to disentangle the complex relationships and interactions among climate, soil, topography and forest structure that can affect species distributions.
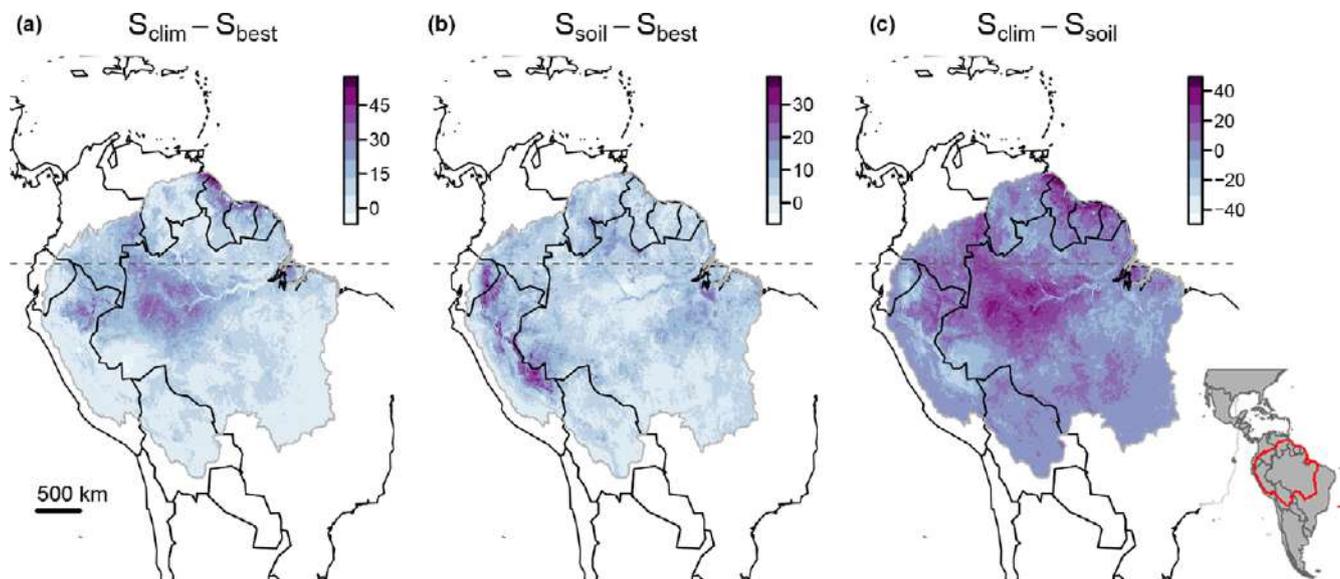
FIGURE 4 Spatial projection of the differences in suitability scores between the SDMs, mapped only in areas defined as unsuitable to species occurrence (averages across 42 plant species). Suitability scores range from 0 (low probability of species occurrence) to 100 (high probability of occurrence). a) The difference in suitability between the CLIM and best models ($S_{clim} - S_{best}$). b). The difference in suitability between the SOIL and best models ($S_{soil} - S_{best}$). c) The difference in suitability between the CLIM and SOIL models ($S_{clim} - S_{soil}$). High values in panels a) and b) denote regions where the CLIM and SOIL models over-predict the suitable conditions relative to the best models respectively. In panel c) values close to zero indicate areas where the CLIM and SOIL models jointly predict edaphic and climatic restrictions to species occurrence; negative values denote areas where the CLIM models predict less suitable condition, but the SOIL models over-predict and the positive values means the opposite. So negative values indicate where climate is the main restricting factor and positive values indicate where soil is the main restricting factor. Maps are in Mollweide equal-area projection, the dashed line denotes the Equator and inset shows the Amazonian limits in red [Colour figure can be viewed at wileyonlinelibrary.com]

Current hydrological trends and recent projections of global climate models predict increasing dry season length for the southern part of Amazonia (Boisier et al., 2015). Increasingly dry conditions may cause forests to become shorter and more open (Hutyra et al., 2005; Levine et al., 2016; Olivares et al., 2015). The species response curves along annual precipitation, dry season length and forest structure gradients suggest that most of our focal species avoid dry conditions and prefer tall and closed forests. This is consistent with the prevailing view that many Amazonian plant species are sensitive to drought (Nepstad, Tohver, Ray, Moutinho, & Cardinot, 2007; Phillips et al., 2010) and that their ranges may contract if dry seasons become more severe (Feeley et al., 2012; Olivares et al., 2015). Temperature showed weak predictive power in our SDMs, suggesting that changes in temperature may be less important for range shifts than changes in water availability.

Massive species loss in Amazonia is expected under climate change if species are not able to either track suitable climate by migration or adapt to the new conditions (Feeley et al., 2012). If species distributions are strongly controlled by soils, species migrations may be effectively prevented by barriers created by large areas of unsuitable soil (Tuomisto et al., 2016). In our analyses, climatic variables alone explained species range limits in fewer areas than soil variables alone did, which suggests that species may indeed face problems in finding suitable routes to track climate change. As soil properties change much slower than climate does, and climate does not seem to be an important driver of soil features in Amazonian

lowlands (Quesada et al., 2010), the presence of edaphic constraints on species ranges can decouple potential range shifts from pure climate tracking.

Disregarding soil variables in fine-scale dynamic models has led to inaccurate predictions of species potential distribution areas under climate change (Bertrand, Perez, & Gégout, 2012). This drawback arises because landscape heterogeneity and metapopulation dynamics limit the ability of species to disperse and track suitable climate changes (Opdam & Wascher, 2004; Schurr et al., 2012). Although the fine spatial configuration of soil properties is not completely reflected in the coarse spatial resolution of our analyses, our results suggest that suitable areas for species occurrence are more patchy than the rather continuous areas that are predicted by climate-only models. The inclusion of edaphic dimension in the forthcoming process-based dynamic models should guide useful conservation and management strategies to protect Amazonian plant species under climate change.

Although the quality of the SoilGrids data layers has not yet been properly assessed, inaccuracies must be expected simply because the density of field sampling in Amazonia is very low and other ecophysiological relevant soil variables, such as concentration of base cations (Ca, Mg, K and Na) and phosphorus are missing in the available soil data. These macronutrients have been shown to be important correlates of floristic patterns and forests structure (Cámara-Leret, Tuomisto, Ruokolainen, Balslev, & Kristiansen, 2017; Figueiredo, Costa, Nelson, & Pimentel, 2014; Higgins et al., 2011;

Quesada et al., 2012; Tuomisto, Ruokolainen, & Yli-Halla, 2003, Tuomisto et al., 2016; Zuquim et al., 2014). CEC is available, but it did not come out as an important variable in our SDMs. The drawback with CEC provided by SoiGrids is that cations are extracted raising the soil pH to 7, creating artificial charges on soil colloids and inaccurate estimates of plant exchangeable cations (Quesada et al., 2011). Therefore, CEC obtained at pH 7 have weak relations with other proxies of soil fertility and should be used only to soil classification purposes (Quesada et al., 2011). In addition, our analyses were carried out at a 5-km resolution, which reduced our ability to capture fine-scale soil variation. A probable consequence of this is that models predict wider edaphic niches than they are in reality, which lead to spatial over-predictions of suitable conditions. It is beyond the scope of the present study to assess the extent of such problems, although they are likely to lead to underestimation of the importance of soil factors in the SDMs.

The relative importance of climate as opposed to edaphic drivers of species distribution has traditionally been considered a matter of scale of analysis. Climate is expected to be useful at broad spatial extent (e.g. global, continental and regional) and soil is expected to be relevant only at small spatial extents and in high-resolution studies (e.g. landscape and local scales) (Eiserhardt et al., 2011; Pearson & Dawson, 2003). In our analyses, soil variables emerged as the most important even at the continental extent. This is no doubt related to relative homogeneity of the climate within Amazonia, at least when compared to subtropical, temperate and montane areas where climatic differences are much larger, and therefore, more restrictive to plant distributions. However, even at global-scale, edaphic variables can better explain ecological patterns (e.g. the variation in some leaf functional traits) than climatic variables (Maire et al., 2015). In Amazonia, soils mediate several ecological patterns and processes, such as large-scale floristic turnover (ter Steege et al., 2006; Tuomisto et al., 2016), biotic interactions (Fine, Mesones, & Coley, 2004), species diversification (Fine, Daly, Muñoz, Mesones, & Cameron, 2005), forest dynamics and resilience (Levine et al., 2016; Quesada et al., 2012). We believe that it is time to scale-up the importance of soil in ecology and biogeography. The inclusion of the soil component in macroecological models will also provide a better framework towards understanding the complex impacts of environmental change on ecological systems.

ORCID

Fernando O. G. Figueiredo http://orcid.org/0000-0002-9333-0708

REFERENCES

Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3, 177–187.

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species – climate impact models under climate change. *Global Change Biology*, 11, 1504–1513.

Baddeley, A., & Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12, 1–42. Retrieved from http://www.jstatsoft.org/v12/i06/

Bertrand, R., Perez, V., & Gégout, J.-C. (2012). Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: The case of Quercus pubescens in France. *Global Change Biology*, 18, 2648–2660.

Blach-Overgaard, A., Svenning, J. C., Dransfield, J., Greve, M., & Balslev, H. (2010). Determinants of palm species distributions across Africa: The relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography*, 33, 380–391.

Boisier, J. P., Ciais, P., Ducharne, A., & Guimberteau, M. (2015). Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Climate Changes*, 5, 656–660.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, 1145–1159.

Buermann, W., Saatchi, S., Smith, T., Zutta, B., Chaves, J., Milá, B., & Graham, C. (2008). Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, 35, 1160–1176.

Burnham, K. K. P., & Anderson, D. R. D. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag.

Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H., & Kristiansen, S. M. (2017). Modelling responses of western Amazonian palms to soil nutrients. *Journal of Ecology*, 105, 367–381.

Chamberlain, S., Ram, K., & Hart, T. (2016). spocc: Interface to species occurrence data sources. R package version 0.5.0. Retrieved from https://CRAN.R-project.org/package=spocc

Costa, R. C. F., Magnusson, W. E., & Luizão, R. C. (2005). Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds. *Journal of Ecology*, 93, 863–878.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1 km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6, 247–254.

Eiserhardt, W., Svenning, J., Kissling, W., & Balslev, H. (2011). Geographical ecology of the palms (Arecaceae): Determinants of diversity and distributions across spatial scales. *Annals of Botany*, 108, 1391–1416.

Ekstrøm, C. (2016). MESS: Miscellaneous Esoteric Statistical Scripts. R package version 0.4-3. Retrieved from https://CRAN.R-project.org/package=MESS

Eva, H. D., & Huber, O. (2005). *A proposal for defining the geographical boundaries of Amazonia*. Luxemburg: Office for Official Publications of the European Communities.

Feeley, K.J. (2012). Distributional migrations, expansions, and contractions of tropical plant species as revealed in dated herbarium records. *Global Change Biology*, 18, 1335–1341.

Feeley, K. J., Malhi, Y., Zelazowski, P., & Silman, M. R. (2012). The relative importance of deforestation, precipitation change, and temperature sensitivity in determining the future distributions and diversity of Amazonian plant species. *Global Change Biology*, 18, 2636–2647.

Figueiredo, F. O. G., Costa, F. R. C., Nelson, B. W., & Pimentel, T. P. (2014). Validating forest types based on geological and land-form features in central Amazonia. *Journal of Vegetation Science*, 25, 198–212.

Fine, P., Daly, D., Muñoz, G., Mesones, I., & Cameron, K. (2005). The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae in the western Amazon. *Evolution*, 59, 1464–1478.

Fine, P., Mesones, I., & Coley, P. D. (2004). Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, 305, 663–665.

Fittkau, E.J., Junk, W., Klinge, H., & Sioli, H. (1975) Substrate and vegetation in the Amazon region. *Berichte der Internationalen Symposien der Internationalen Vereinigung für Vegetationskunde Herausgegeben von Reinhold Tüxen* (ed. by J. Cramer), pp. 73–90. J. Cramer, Vaduz.

Gaston, K. J. (2003). *The structure and dynamics of geographical ranges*. Oxford: Oxford University Press.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. (2008). Weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistic*, 2, 1360–1383.

Gelman, A, & Su, Y-S. (2016). arm: Data analysis using regression and multilevel hierarchical models. R package version 1.9-1. Retrieved from https://CRAN.R-project.org/package=arm

Grubb, P. T. (1977). Control of forest growth and distribution on wet tropical mountains: With special reference to mineral nutrition. *Annual Review of Ecology and Systematics*, 8, 83–107.

Hayes, F. E., & Sewlal, J. A. N. (2004). The Amazon River as a dispersal barrier to passerine birds: Effects of river width, habitat and taxonomy. *Journal of Biogeography*, 31, 1809–1818.

Hefley, T. J., & Hooten, M. B. (2015). On the existence of maximum likelihood estimates for presence-only data. *Methods in Ecology and Evolution*, 6, 648–655.

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., ... Gonzalez, M. R. (2014). SoilGrids1 km—Global soil information based on automated mapping. *PLoS ONE*, 9, e105992.

Higgins, M., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O. L., ... Räsänen, M. (2011). Geological control of floristic composition in Amazonian forests. *Journal of Biogeography*, 38, 2136–2149.

Hijmans, J.R. (2016). raster: Geographic data analysis and modeling. R package version 2.5-8. Retrieved from https://cran.r-project.org/web/packages/raster/.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 1978, 1965–1978.

Hijmans, R.J., Phillips, S., Leathwick, J., & Elith, J. (2017). dismo: Species Distribution Modeling. R package version 1.1-4. Retrieved from https://CRAN.R-project.org/package=dismo

Hutyra, L. R., Munger, J. W., Nobre, C. A., Saleska, S. R., Vieira, S. A., & Wofsy, S. C. (2005). Climatic variability and vegetation vulnerability in Amazônia. *Geophysical Research Letters*, 32. https://doi.org/10.1029/2005GL024981

Irion, G. (1978). Soil infertility in the Amazonian rain forest. *Naturwissenschaften*, 10, 515–519.

Jenny, H. (1994). *Factors of soil formation: A system of quantitative pedology*. New York: Dover Publications.

Kempes, C. P., West, G. B., Crowell, K., & Girvan, M. (2011). Predicting maximum tree heights and other traits from allometric scaling and resource limitations. *PLoS ONE*, 6, e20551.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., & Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15, 809–817.

Levine, N. M., Zhang, K., Longo, M., Baccini, A., Phillips, O. L., Lewis, S. L., ... Moorcroft, P.R. (2016). Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences*, 113, 793–797.

Maire, V., Wright, I. J., Prentice, I. C., Batjes, N. H., Bhaskar, R., van Bodegom, P. M., ... Santiago, L. S. (2015). Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecology and Biogeography*, 24, 706–717.

Manel, S., Williams, H. C., & Ormerod, S. J. (2001). Evaluating presence absence models in ecology; the need to count for prevalence. *Journal of Applied Ecology*, 38, 921–931.

Marage, D., & Gégout, J.-C. (2009). Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, 18, 88–97.

Merow, C., Smith, M. J. & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, 36, 1058–1069.

Miles, L., Grainger, A., & Phillips, O. (2004). The impact of global climate change on tropical forest biodiversity in Amazonia. *Global Ecology and Biogeography*, 13, 553–565.

Moscoso, V., Albernaz, A. L., & Salomão, R. P. (2013). Niche modelling for twelve plant species (six timber species and six palm trees) in the Amazon region, using collection and field survey data. *Forests Ecology and Management*, 310, 652–662.

Naimi, B. (2015). usdm: Uncertainty analysis for species distribution models. R package version 1.1-15. Retrieved from https://CRAN.R-project.org/package=usdm

Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P., & Cardinot, G. (2007). Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology*, 88, 2259–2269.

Olivares, I., Svenning, J.-C., van Bodegom, P. M., & Balslev, H. (2015). Effects of warming and drought on the vegetation and plant diversity in the Amazon Basin. *The Botanical Review*, 81, 42–69.

Opdam, P., & Wascher, D. (2004). Climate change meets habitat fragmentation: Linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, 117, 285–297.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12, 361–371.

Pebesma, E.J, & Bvand, R.S. (2005). Classes and methods for spatial data in R. R News 5(2). Retrieved from https://cran.r-project.org/doc/Rnews/

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.

Phillips, O. L., van der Heijden, G., Lewis, S. L., Lopez-Gonzales, G., Aragão, L. E. O. C., Lloyd, J., ... Vilanova, E. (2010). Drought-mortality relationships for tropical forests. *The New Phytologist*, 187, 631–46.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.

Quesada, C.A., Lloyd, J., Anderson, L.O., Fyllas, N.M, Schwarz, M., & Czimczick, C.I. (2011). Soils of Amazonia with particular references to the RAINFOR sites. *Biogeosciences*, 8, 1415–1440.

Quesada, C. A., Lloyd, J., Schwarz, M., Patiño, S., Baker, T. R., Czimczik, C., ... Paiva, R. (2010). Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences*, 7, 1515–1541.

Quesada, C. A., Phillips, O. L., Schwarz, M., Czimczik, C. I., Baker, T. R., Patiño, S., ... Lloyd, J. (2012). Basin-wide variations in Amazon forest

structure and function are mediated by both soils and climate. *Biogeosciences*, 9, 2203–2246.

R Development Core Team (2015). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Räsänen, M. E., Salo, J. S., & Kalliola, R. J. (1987). Fluvial perturbance in the western Amazon basin: Regulation by long-term sub-Andean tectonics. *Science*, 238, 1398–1401.

Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S., ... Warton, D. I. (2015). Point process models for presence-only analysis —a review. *Methods in Ecology and Evolution*, 6, 366–379.

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69, 274–281.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, 54, 507–554.

Robertson, M. (2016). biogeo: Point data quality assessment and coordinate conversion. R package version 1.0. Retrieved from https://CRAN.R-project.org/package=biogeo

Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., ... Zimmermann, N. E. (2012). How to understand species' niches and range dynamics: A demographic research agenda for biogeography. *Journal of Biogeography*, 12, 2146–2162.

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116. https://doi.org/10.1029/2011JG001708

Sombroek, W. (2000). Amazon landforms and soils in relation to biological diversity. *Acta Amazonica*, 30, 81–100.

ter Steege, H., Pitman, N. C. A., Phillips, O. L., Chave, J., Sabatier, D., Duque, A., ... Vásquez, R. (2006). Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, 443, 444–447.

Thomas, E., Caicefo, C. A., Loo, J., & Roeland, K. (2014). The distribution of the Brazil nut (*Bertholletia excelsa*) through time: From range contraction in glacial refugia, over human-mediated expansion, to anthropogenic climate change. *Boletim do Museu Paraense Emílio Goeldi. Ciências Naturais*, 9, 267–291.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., & van Etten, J. (2012). Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. *PLoS ONE*, 7, e47676.

Thuiller, W. (2007). Climate change and the ecologist. *Nature*, 448, 550–552.

Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Chuviña, J., ... Bongers, F. (2011). Patterns and determinants of floristic variation across lowland forests of Bolivia. *Biotropica*, 43, 405–413.

Tuomisto, H., Moulatlet, G., Balslev, H., Emilio, T., Figueiredo, F. O. G., Pedersen, D., & Ruokolainen, K. (2016). A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *Journal of Biogeography*, 12, 2400–2411.

Tuomisto, H., & Poulsen, A. D. (1996). Influence of edaphic specialization on pteridophyte distribution in neotropical rain forests. *Journal of Biogeography*, 23, 283–293.

Tuomisto, H., Ruokolainen, K., Aguilar, M., & Sarmiento, A. (2003). Floristic patterns along a 43-km long transect in an Amazonian rain forest. *Journal of Ecology*, 91, 743–756.

Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, 299, 241–244.

Vedel-Sørensen, M., Tovaranonte, J., Bøcher, P., Balslev, H., & Barfod, A. (2013). Spatial distribution and environmental preferences of 10 economically important forest palms in western South America. *Forest Ecology and Management*, 307, 284–292.

Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics*, 4, 1383–1402.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., ... Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, 88, 15–30.

Zuquim, G., Tuomisto, H., Jones, M. M., Prado, J., Figueiredo, F. O. G., Moulatlet, G. M., ... Emilio, T. (2014). Predicting environmental gradients with fern species composition in Brazilian Amazonia. *Journal of Vegetation Science*, 25, 1195–1207.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

BIOSKETCH

Author contributions: F.F and F.C. conceived the idea and led the writing; F.F. analysed the data; all authors collected data and helped with the writing.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.