Received: 30 November 2018 | Revised: 1 March 2019 | Accepted: 5 April 2019

DOI: 10.1111/jbi.13627

SPECIAL ISSUE

Journal of Biogeography | WILEY

Discovering floristic and geoecological gradients across Amazonia

Hanna Tuomisto[1] | Jasper Van doninck[1,2] | Kalle Ruokolainen[1] | Gabriel M. Moulatlet[1,3] | Fernando O. G. Figueiredo[4] | Anders Sirén[1] | Glenda Cárdenas[1] | Samuli Lehtonen[5] | Gabriela Zuquim[1]

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

[2]Department of Geography and Geology, University of Turku, Turku, Finland

[3]Universidad Regional Amazónica Ikiam, Tena, Ecuador

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

[5]Biodiversity Unit, University of Turku, Turku, Finland

Correspondence
Hanna Tuomisto, Department of Biology, University of Turku, 20014 Turku, Finland.
Email: hanna.tuomisto@utu.fi

Funding information
Academy of Finland, Grant/Award Number: 139959, 273737 and 296406; Suomen Kulttuurirahasto; FAPEAM; CNPq; University of Turku

Editor: Simon Scheiter

Abstract

Aim: To map and interpret floristic and geoecological patterns across the Amazon basin by combining extensive field data with basin-wide Landsat imagery and climatic data.

Location: Amazonia.

Taxon: Ground truth data on ferns and lycophytes; remote sensing results reflect forest canopy properties.

Methods: We used field plot data to assess main ecological gradients across Amazonia and to relate floristic ordination axes to soil base cation concentration, Climatologies at High Resolution for the Earth's Land Surface Areas (CHELSA) climatic variables and reflectance values from a basin-wide Landsat image composite with generalized linear models. Ordination axes were then predicted across all Amazonia using Landsat and CHELSA, and a regional subdivision was obtained using k-medoid classification.

Results: The primary floristic gradient was strongly related to base cation concentration in the soil, and the secondary gradient to climatic variables. The Landsat image composite revealed a tapestry of broad-scale variation in canopy reflectance characteristics across Amazonia. Ordination axis scores predicted using Landsat and CHELSA variables produced spatial patterns consistent with existing knowledge on soils, geology and vegetation, but also suggested new floristic patterns. The clearest dichotomy was between central Amazonia and the peripheral areas, and the available data supported a classification into at least eight subregions.

Main conclusions: Landsat data are capable of predicting soil-related species compositional patterns of understorey ferns and lycophytes across the Amazon basin with surprisingly high accuracy. Although the exact floristic relationships may differ among plant groups, the observed ecological gradients must be relevant for other plants as well, since surface reflectance recorded by satellites is mostly influenced by the tree canopy. This opens exciting prospects for species distribution modelling, conservation planning, and biogeographical and ecological studies on Amazonian biota. Our maps provide a preliminary geoecological subdivision of Amazonia that can now be tested and refined using field data of other plant groups and from hitherto unsampled areas.

KEYWORDS
Amazonia, biogeographical regions, floristic gradients, Landsat satellite, medium resolution multispectral imagery, remote sensing, tropical rain forest, vegetation

1 | INTRODUCTION

Amazonian rainforests have long fascinated biologists, but their internal heterogeneity remains poorly understood. One of the open questions is how to best subdivide them into floristically and biogeographically meaningful units. Such subdivision would be useful for biogeographical inferences, conservation planning, and forecasting potential effects of climate change or deforestation on species distributions and their viability.

Amazonia has been subdivided into areas of endemism using information about species and subspecies distributions especially for birds (Cracraft, 1985; Haffer, 1974; Ribas, Aleixo, Nogueira, Miyaki, & Cracraft, 2012), but the low collecting density within Amazonia necessarily renders maps based on species occurrence records speculative. The most detailed subdivision of Amazonia is probably the WWF ecoregion map, which takes into account both landscape features and information about rivers as dispersal barriers (Olson et al., 2001). Animal distributions and abundances have been found to vary between seasonally inundated versus non-inundated forests and even in response to more subtle changes in soil productivity and floristic composition (Halme & Bodmer, 2007; Peres, 1999; Stevenson, 2014). Therefore, knowledge on floristic variability and related environmental heterogeneity can provide the kind of baseline information needed to understand general biogeographical relationships and ecological patterns across Amazonia.

An early geochemical subdivision of Amazonia (Fittkau, Junk, Klinge, & Sioli, 1975) recognized four major regions: central Amazonia with poor sedimentary soils, western Amazonia with richer soils derived from Andean material, and northern and southern peripheral regions on the Precambrian formations of the Guyanan and Brazilian shields, respectively. A somewhat more detailed classification was used by ter Steege et al. (2013), and Sombroek (2000) drafted a classification focusing on soil types. In these cases, the authors recognized the importance of geochemical differences for vegetation or species richness but did not elaborate on how floristic composition might have varied among the recognized regions.

Floristic studies in the past few decades have shown that species compositional patterns of many plant groups in Amazonia, including both trees and understorey plants, are linked to soil properties (Baldeck, Tupayachi, Sinca, Jaramillo, & Asner, 2016; Duivenvoorden, 1995; Higgins et al., 2012, 2011; Pansonato, Costa, de Castilho, Carvalho, & Zuquim, 2013; Phillips et al., 2003; Ruokolainen, Tuomisto, Macía, Higgins, & Yli-Halla, 2007; Tuomisto et al., 2016; Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, & Sarmiento, 2003; Tuomisto, Ruokolainen, & Yli-Halla, 2003). Soil properties, in turn, depend to a large degree on the soil parent material, especially its mineralogy, sedimentation history and the time it has been exposed to weathering. Therefore, considering the complex geological history of Amazonia is relevant when studying species distributions and floristic variability. Geological formations in Amazonia have had various origins. There are, for example, Precambrian cratonic shields, Miocene semimarine to lacustrine deposits, recent deposits of material from Andean volcanic eruptions and sediments consisting of material that has been leached to various degrees during cycles of fluvial sedimentation and resuspension (Fittkau et al., 1975; Hoorn, 1993; Hoorn et al., 2010; Räsänen, Linna, Santos, & Negri, 1995; Räsänen, Neller, Salo, & Jungner, 1992; Sombroek, 2000). As a result, Amazonian soils have become highly heterogeneous, and the concentrations of major plant nutrients in them vary over more than two orders of magnitude (Hengl et al., 2014; Sanchez & Buol, 1974; Tuomisto et al., 2016; Tuomisto, Ruokolainen, & Yli-Halla, 2003).

Given the huge spatial extent of Amazonia, any mapping effort is faced with practical difficulties, especially data scarcity. Some studies on forest structural properties and tree species richness have applied simple spatial interpolation techniques to cover areas between field sampling points (Stropp, ter Steege, Malhi, ATDN, & RAINFOR, 2009; ter Steege et al., 2006, 2003). Many others have taken advantage of the possibilities offered by remote sensing. For example, data from coarse-resolution sensors have been used to assess biomass, productivity and seasonality in leaf production across all of Amazonia (Mitchard et al., 2014; Saatchi, Houghton, Dos Santos Alvalá, Soares, & Yu, 2007; Saatchi et al., 2009; Wagner et al., 2017). Vegetation mapping has taken advantage of radar images and aerial photographs to identify forest types that differ in structural and terrain characteristics (Duivenvoorden & Lips, 1993; Huber & Alarcón, 1988; Huber, Gharbarran, & Funk, 1995; IBGE, 2004). Several studies have used Landsat data to predict edaphic properties or different aspects of plant communities (species composition, turnover or richness) over landscape extents (Draper et al., 2019; Higgins et al., 2012, 2011; Salovaara, Thessler, Malik, & Tuomisto, 2005; Sirén, Tuomisto, & Navarrete, 2013; Thessler, Ruokolainen, Tuomisto, & Tomppo, 2005; Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003). The highest local resolution has been obtained by airborne hyperspectral sensors and LiDAR, which have been used to map forest properties such as canopy height, above-ground carbon stocks and canopy chemistry at the regional extent (Asner et al., 2015, 2013, 2014). However, floristic patterns at the basin-wide extent remain to be clarified.

Here we have two main objectives. Firstly, we document the main floristic gradients across the Amazon basin and their environmental correlates. Secondly, we assess to what degree these floristic and environmental patterns can be identified and mapped across the basin. We address these questions using a combination

of an extensive field dataset (1,572 plots with data on ferns, lycophytes and soils), climatic data layers and a basin-wide composite of Landsat imagery. Finally, we draw some ecological and biogeographical conclusions from the results and derive a floristic subdivision of Amazonia to stimulate further research on this topic.

2 | MATERIAL AND METHODS

2.1 | Floristic, environmental and remotely sensed data

We combined floristic and soil data collected independently by two teams, the Amazon Research Team of the University of Turku (UTU) and the Brazilian Program in Biodiversity Research (PPBio). When put together, these inventories span across a large part of the Amazon basin and cover an elevation range of about 50–600 m.

The field sampling protocols of the two teams were slightly different. The PPBio dataset consists of 309 plots of 2 m by 250 m (500 m^2) that follow the terrain contours to avoid local topographical variation (Magnusson et al. 2005). Of these, 102 originated from four permanent PPBio grids, where the plots were placed at 1-km intervals, and the rest were more scattered. The UTU dataset consists of 388 line transects that were 2 m or 5 m wide and either 500 m, 1,300 m or 9.8–43 km long (Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003). They followed a predetermined compass bearing, thus crossing the local topographical variation. For the purposes of the present paper, we subdivided the UTU transects into contiguous non-overlapping subunits that as closely as possible matched the plot size in the PPBio dataset, i.e. either 5 m by 100 m (500 m^2) or 2 m by 200 m (400 m^2). We retained those 1,263 transect subunits (plots) in which at least one soil sample had been collected. The 500-m-transects yielded 1–3 plots each (374 transects, 1,015 plots), the 1,300-m-transects 1–13 plots each (8 transects, 68 plots) and the longer transects about 1 plot for each km of transect length (6 transects, 180 plots). In total, 1,572 plots from 697 separate transects were available for analysis. The vast majority of these represent non-inundated terra firme forests, but some were situated in intervening swamps or floodplains of small rivers. All plots were georeferenced using coordinates obtained in the field with hand-held GPS devices.

Within each of the 1,572 plots, terrestrial fern and lycophyte individuals that had at least one leaf (leafy stem in the case of lycophytes) longer than 10 cm were recorded. Species that are mostly epiphytic or hemiepiphytic were excluded. We focused on ferns and lycophytes because these are of moderate size, relatively abundant and not too species-rich, which made it possible to obtain floristically representative samples from many sites. At local to landscape extents, their species turnover patterns have been found to closely mirror both species turnover patterns in other plants (including trees) and differences in soils (Duque et al., 2005; Higgins et al., 2011; Jones et al., 2013; Pansonato et al., 2013; Ruokolainen et al., 2007; Tuomisto et al., 2016; Vormisto, Phillips, Ruokolainen, Tuomisto, & Vásquez, 2000). Since ferns and lycophytes have high dispersal ability, they may indicate geoecological patterns more reliably but suggest less biogeographical differentiation across Amazonia than more dispersal-limited plants would (Tuomisto, Ruokolainen, & Yli-Halla, 2003).

Voucher specimens of each species were collected during each field campaign for verification of species identifications. Vouchers are deposited in one or two herbaria in the country of origin (AMAZ, USM and/or CUZ in Peru, QCA and QCNE in Ecuador, COAH in Colombia, SP and/or INPA in Brazil, CAY in French Guiana) and for the UTU data also in TUR (herbarium acronyms according to Thiers, continuously updated). Initial identifications were done for the UTU and PPBio datasets separately, but identifications were harmonized by H.T. and G.Z. on the basis of either the vouchers themselves or photographs of them. Several species pairs or species complexes were so similar that they could not always be reliably separated in the field. These were lumped to ensure a consistent taxonomy over the entire dataset (for simplicity, both true species and species complexes will be referred to as species).

Composite samples of the surface soil (top 5 or 10 cm after removing the litter layer, 5–6 subsamples) were collected within each plot for chemical and textural analyses. In the PPBio plots, the subsamples were taken at 50-m intervals along the long axis of the plot. In the UTU plots, each soil sample was taken within an area of about 5 m by 5 m, but the samples collectively represented different topographical positions within a transect. Further details on the field and laboratory methodology are available in earlier studies (Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003; Tuomisto et al., 2017; Zuquim et al., 2014). When two soil samples were available for the same UTU plot, we averaged their measurement values for the analyses. Here, we focus on the sum of the concentrations of exchangeable Ca, Mg and K (measured in cmol(+)/kg), which we will refer to as base cation concentration. All analyses were carried out using logarithmically transformed values (logCat). This variable was chosen because earlier studies have found its relationship with plant species composition and changes therein to be consistently strong, even when the relationships with the individual cations have varied among regions (Higgins et al., 2011; Ruokolainen et al., 2007; Tuomisto et al., 2016; Tuomisto, Poulsen, et al., 2003; Zuquim et al., 2014).

Although digital soil maps exist (Hengl et al., 2017), they do not provide layers of soil base cation concentration. The digitally available variables, such as soil type and cation exchange capacity have had only weak relationships with field-measured base cation concentration (Moulatlet et al., 2017). Therefore, we used Landsat imagery as a surrogate, since earlier studies at the landscape extent have found them useful for identifying spatial heterogeneity in soils and soil-related floristic patterns (Higgins et al., 2012, 2011; Salovaara et al., 2005; Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003).

Landsat image analyses across all Amazonia were made possible by a recent basin-wide composite of Landsat TM/ETM+ images. The composite is based on more than 16,000 sufficiently cloud-free acquisitions from the 10-year period 2000–2009, and it has already been shown to predict soil base cation concentration relatively

well (Van doninck & Tuomisto, 2018). Technical details on how the image composite was produced have been described elsewhere (Van doninck & Tuomisto, 2017a,2017b,2018). Climatic data (19 bioclimatic variables) at 30 arcsec resolution (approximately 1 km) were obtained from the Climatologies at High Resolution for the Earth's Land Surface Areas (CHELSA, Karger et al., 2017) (http://chelsa-climate.org/).

2.2 | Data analyses

To identify floristic gradients, we first calculated compositional dissimilarities among the plots using the Sørensen index, which is the presence-absence version of the Bray-Curtis index. Almost 30% of the dissimilarity values were saturated to the maximum value of unity, so we used the extended (step across) version of the index in order to obtain ecologically realistic dissimilarities between the plots that shared no species (De'ath, 1999; Tuomisto, Ruokolainen, & Ruokolainen, 2012). Then we performed an ordination based on Principal Coordinates Analysis (PCoA) to visualize the main floristic gradients. We used univariate linear regression analysis to assess the degree to which each of the first three PCoA axes (PCoA 1–3) were related to the explanatory variables (log-transformed soil base cation concentration, bioclimatic variables and reflectance values in the Landsat TM/ETM+ composite).

Landsat bands 1 (blue) and 2 (green) were not used because they were very noisy due to residual atmospheric contamination. For the remaining visible band 3 (red) and the near to shortwave infrared bands 4, 5 and 7, we extracted the reflectance values corresponding to the coordinates of each plot in three different ways: (a) full-resolution data at 1 arcsec (approximately 30 m) resolution, (b) data retaining the full resolution but filtered by passing a moving widow of 15 by 15 pixels over the image and assigning to each pixel the median value from the window centred on it, and (c) low resolution data obtained by coarsening the resolution to pixels of 15 arcsec (approximately 450 m) by assigning to each new pixel the median value from the corresponding 15 by 15 original pixels. Non-forested pixels were masked using an unsupervised k-means clustering with post-classification interpretation based on visual inspection of spectral signature and spatial distribution. For (a) and (b), masking was based on the original full-resolution data, for (c) it was based on the coarsened pixels.

Many of the CHELSA variables were mutually highly correlated. To keep a reduced but representative set of climate variables for modelling, we calculated the variance inflation factor (VIF) for all 19 variables. Then we iteratively excluded the variable with the highest VIF and recalculated VIF for the remaining variables until none of the VIF values exceeded 50. All analyses were then based on the remaining eleven CHELSA variables. These represented temperature variability (Bio 2–4), mean temperature (Bio 8, 9, 11), mean precipitation (Bio 13, 14, 18, 19) and precipitation variability (Bio 15).

To formally model floristic and soil gradients across the Amazon basin, we constructed generalized linear models (GLMs) with each of the PCoA axes 1–3 and logCat as the response variable in turn. GLMs using different combinations of predictor variables were tested by randomly dividing the data into ten folds and using each fold as an independent test set in turn. Adjacent plots that were part of the same field transect were allowed to go to separate folds only if the distance between them was at least 1 km. We also evaluated elevation from SRTM digital elevation model and textural data layers obtained from standard deviations within the window of Landsat pixels as predictive variables. GLMs using the Landsat median values, CHELSA variables and both together were applied over the entire Amazon basin in order to produce predictive maps of the main floristic gradients (PCoA axes 1–3). The predictive maps were then classified using an implementation of k-medoids clustering for large applications (CLARA).

All analyses were carried out in R version 3.4.1 (R Core Team, 2017) using the packages "raster" for raster image analysis (Hijmans, 2017), "vegan" for ordinations (Oksanen et al., 2017), "cluster" for unsupervised clustering (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2016), and "stats" for GLM (R Core Team, 2017).

3 | RESULTS

3.1 | Amazonian heterogeneity as seen from space

The Landsat TM/ETM+ colour composite reveals a tapestry of broad-scale variation in canopy reflectance characteristics across Amazonia (Figure 1; unannotated version can be found in Appendix S1: Figure S1a in Supporting Information). The parallel Landsat flight paths remain to some degree visible in the red band 3 (assigned to the blue colour channel in Figure 1 and Figure S1a). This is because correction of atmospheric effects in the image composite was not perfect, and the shorter visible wavelengths are more affected by scattering caused by aerosols than the longer infrared ones. The false colour composite based on the infrared bands 4, 5 and 7 is virtually seamless (Figure S1b). Nevertheless, adding a visible wavelength improves the separability of nuances, as the infrared bands are highly correlated with each other (Appendix S1: Figure S2).

Several known geo-ecological entities of floristic relevance can be recognized in the raw Landsat composite. Some of the most notable ones correspond to traditionally recognized vegetation formations, including the heterogeneous floodplain forests along major rivers, swamp forests like those in the Pastaza fan ("Pas" in Figure 1; Räsänen et al., 1992), bamboo-dominated forests like those in Acre ("Bam"; Carvalho et al., 2013), and white-sand forests like those in the upper Rio Negro basin ("WS"; Adeney, Christensen, Vicentini, & Cohn-Haft, 2016). Other recognizable features include the floristic turnover zone that has been suggested to correspond to the limit between the Solimões and Içá formations ("S" and "I", respectively; Higgins et al., 2011; Schobbenhaus et al., 2004; Tuomisto et al., 2016), the Brazilian and Guyanan Precambrian shields ("Bra" and "Gui", respectively) and sandy megafan formations in Roraima ("MF"; Rossetti, Cassola Molina, & Cremon, 2016).

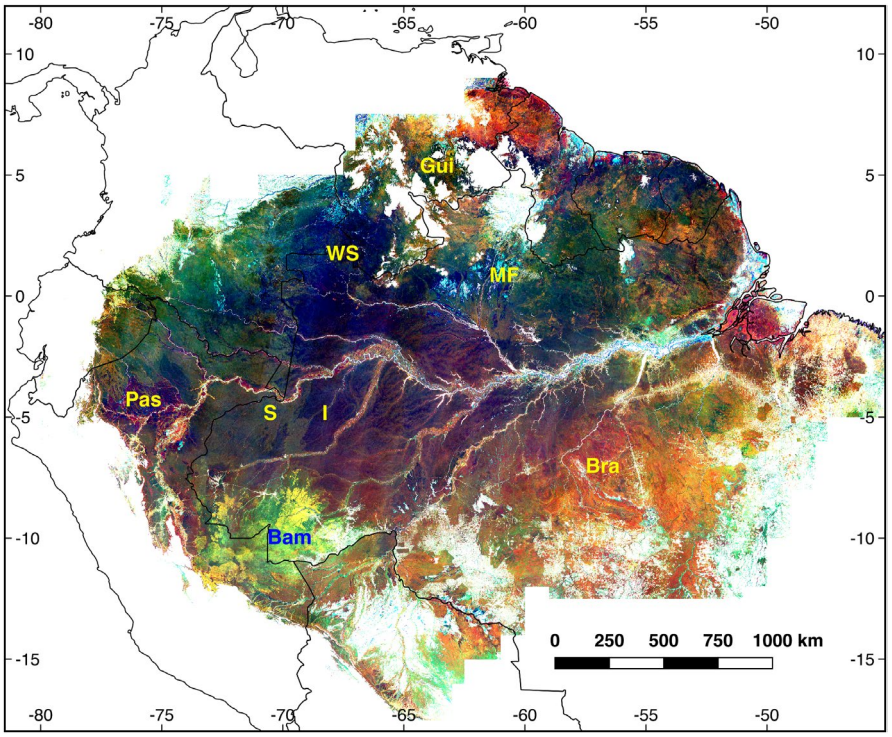

FIGURE 1 Pixel-based Landsat TM/ETM+ composite over Amazonia, based on all Landsat images acquired in July–September during the 10-year period 2000–2009. For this quick view image, the original 30-m image resolution has been coarsened by gridding to 450 m, with each grid cell given the median reflectance value of each band from the 15 by 15 pixels corresponding to the grid cell. Red, green and blue have been assigned, respectively, to bands 4, 7 and 3. Areas that were either classified as non-forest on the basis of the reflectance data or are above 600 m elevation have been masked out and appear white. Pas: Pastaza fan; Bam: bamboo forest, S: Solimões Formation; I: Içá Formation; WS: white sands; MF: megafans; Gui: Guiana shield; Bra: Brazilian shield. Versions without the elevation limit and free of the annotation of both this colour composite and a colour composite based on bands 4, 5 and 7 are available in Figure S1

3.2 | Floristic gradients and their environmental relationships

To obtain a general idea of the ecological drivers of floristic variation across the Amazon basin (Figure 2a), we first identified the main gradients in fern and lycophyte species composition using PCoA ordination (Figure 2b). The main spatial pattern at the basin-wide extent was that western Amazonia was very heterogeneous in both floristic composition and soil base cation concentration with many plots having high values, whereas most plots from central Amazonia had low values for both PCoA axis 1 and soil base cation concentration.

Indeed, the strongest floristic gradient was tightly related to base cation concentration in the soil: a simple linear regression with logCat as the predictor variable explained 75% of the variation in the ordination scores along PCoA axis 1 (Figure 2c; Table 1). Adding either filtered Landsat spectral data, CHELSA climatic data or both to GLMs that already included soil base cation concentration increased mean cross-validated predictive power only by 4–7% (R^2 = 79–82%). Without soil cation concentration in the model, mean predictive power in cross-validation was 47% with filtered Landsat spectral values, 38% with CHELSA climatic values and 58% with both together (Table 1). Soil base cation concentration was almost as well predicted as the first floristic gradient by the Landsat data, but the contribution of the climatic data was less in the case of soils in both absolute and relative terms. This supports the use of Landsat data as a surrogate for mostly soil-related environmental gradients (Table 1 and Table S1, Figures S3 and S4). Elevation as derived from SRTM and textural data from Landsat had little predictive power (Table S1).

The original Landsat resolution (30 m) gives fine detail that allows visual interpretation of the landscape. For example, Figure 3a clearly shows ridge-swale structures in the river floodplain, creeks in non-inundated terrain and the extents of the Içá Formation (dark green–brown), Solimões Formation (pale bluish gray) and their transition zone (pink–orange). These patterns get blurred when pixels are aggregated by factor 15 (Figure 3c), with the filtered image being intermediate (Figure 3b). However, at the 30-m resolution there is so much local variation in pixel values that the predictive power of GLMs for PCoA axis 1 was clearly lower when based on Landsat data with original 30-m pixels (R^2=21%) than when based on the filtered data (R^2 = 47%; Table 1 and Table S1). When projected on a map, the predictions obtained using the original pixels were clearly noisier and provided a poorer contrast among the floristically and edaphically different kinds of forest than predictions obtained with the filtered or coarsened data did (Figure 3d–i).

The second floristic gradient formed a west-to-east spatial trend. Correspondingly, PCoA axis 2 was related to many bioclimatic variables, with the maximum temperature of the warmest month (Bio5) providing the best explanatory power (R^2 = 42%) in a univariate linear

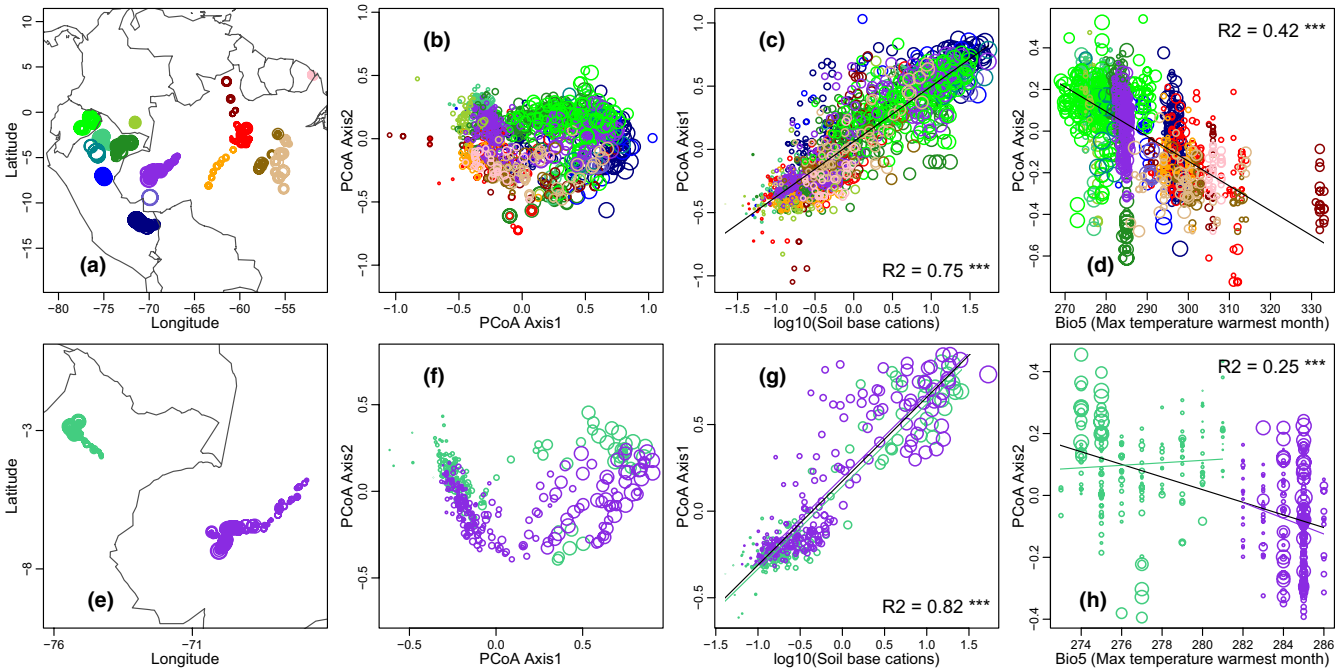

FIGURE 2 (a) Map of 1,572 Amazonian plant inventory plots used in the analyses. (b) Principal Coordinates Analysis (PCoA) ordination based on floristic dissimilarities between the 1,572 plots (presence-absence data of terrestrial ferns and lycophytes, extended Sørensen dissimilarity). (c) Relationship between the first floristic PCoA axis and soil base cation concentration as measured from soil samples taken in each plot. (d) Relationship between the second PCoA axis and maximum temperature of the warmest month as modelled by the CHELSA variable Bio5. (e–h) Same as (a–d), but with all analyses based on a subset of two regions, together containing 398 plots, to show graphs with less cluttering. These regions were selected because they are far apart but have analogous geological settings. In all panels, colours indicate the different geographical regions and symbol size is proportional to log-transformed concentration of exchangeable base cations (Ca, Mg, K) in the soil. The first three axes in the ordination of the full dataset account for 38%, 13% and 8% of the total variation and in the ordination of the two regions for 51%, 11% and 7%

TABLE 1 The degree to which floristic ordination axes (PCoA shown in Figure 2b) and soil base cation concentration can be modelled with data derived from a Landsat TM/ETM+ image composite, the 11 CHELSA climate variables that were sufficiently uncorrelated with each other to have VIF < 50, and logarithmically transformed soil base cation concentration as measured from soil samples taken in each field plot (logCat). Adjusted R^2 (in %) is given for the linear regression between GLM-predicted and measured values. "All R^2" refers to the fit of a GLM model that uses all field plots, "Test R^2" gives the average of R^2 values obtained using cross-validation. Here, the plots were randomly divided into ten folds, each of which was used as an independent validation dataset in turn. Plots from the same transect were kept in the same fold if their distance was less than 1 km. PCoA, Principal Coordinates Analysis; Landsat 30, single pixel values at the field plot locality; Landsat 450, median of 15 by 15 pixel windows centred on the field plot; VIF, Variance Inflation Factor; GLM, Generalized Linear Model

	logCat		PCoA axis 1		PCoA axis 2		PCoA axis 3	
Predictors	All R^2	Test R^2	All R^2	Test R^2	All R^2	Test R^2	All R^2	Test R^2
Landsat 30	15	15	21	21	6	7	3	4
Landsat 450	42	41	48	47	10	11	5	6
CHELSA	30	28	40	38	48	48	17	16
Landsat 450 + CHELSA	49	46	60	58	50	49	18	16
logCat			75	74	0	1	0	2
Landsat 450 + logCat			79	78	10	11	9	9
CHELSA + logCat			80	79	50	50	20	18
Landsat 450 + CHELSA + logCat			83	82	51	50	22	20

regression (Figure 2d; Figure S4). Indeed, GLM models with CHELSA climate variables had mean predictive power of 48% in cross-validation. Adding Landsat variables to the model only increased predictive power marginally (to 49%). This indicates that PCoA axis 2 is mainly a climatic gradient. Landsat variables on their own explained less than 10% of the variation along PCoA axis 2, and soil base cation concentration none at all (Table 1 and Table S1). The third PCoA axis was mostly predicted by the climatic variables, although not very well (R^2 = 16%).

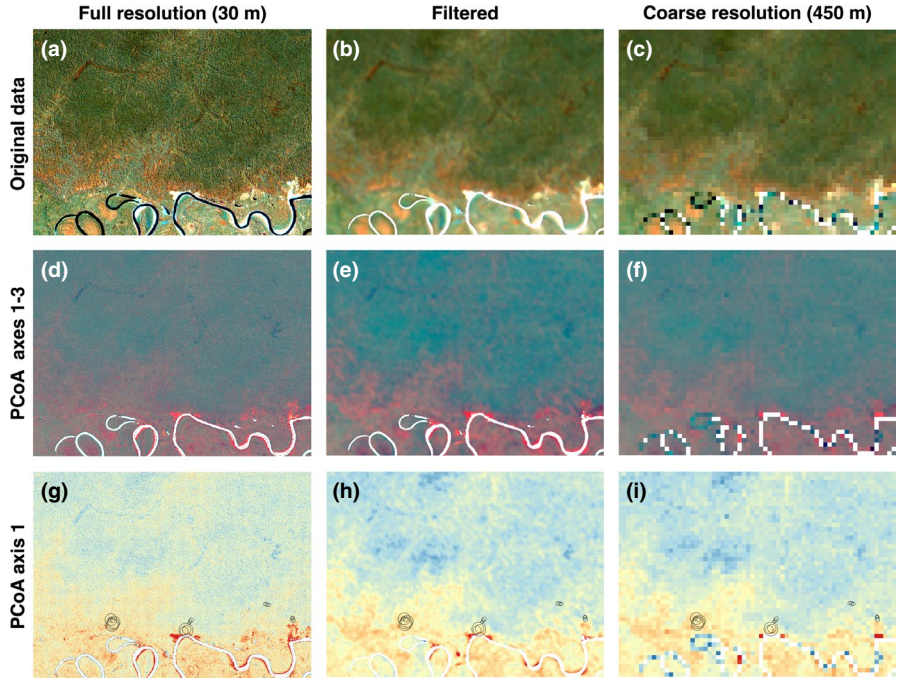

FIGURE 3 Terrain characteristics from a site along the middle Juruá River in Amazonia (midpoint at 68.764°W, 6.400°S). (a–c) False colour composite (bands 4, 5 and 7) of the Landsat TM/ETM+ imagery at full resolution (a), full resolution filtered by assigning each pixel the median value of the 15-by-15-pixel window centred on it (b) and coarsened to 450-m resolution (15-by-15-pixel grid) (c). In the non-inundated areas north of the Juruá floodplain, dark green–brown corresponds to the Içá Formation, pale bluish gray to the Solimões Formation and pink–orange to their transition zone. (d–f) Values predicted for PCoA axes 1–3 on the basis of GLMs trained with the Amazon-wide fern and lycophyte ordination data shown in Figure 2b. Axis 1 is assigned to red, axis 2 to green and axis 3 to blue. Predictions in each panel are based on the reflectance values from the colour composite above it. (g–h) Same as (d–f), but only values predicted for PCoA axis 1 are shown with blue corresponding to low values and red to high values. Each black circle represents one field plot with diameter proportional to log-transformed soil base cation concentration. Colour schemes are the same for all panels on the same row. Panel (c) is a detail of Figure S1b and panel (i) of Figure 4b. GLM, generalized linear model; PCoA, Principal Coordinates Analysis

The environmental relationships of the first two floristic gradients become even clearer when two subsets of the data are compared, one from the Tigre river basin in the west and the other from the Juruá river basin about 1,000 km further south-east (Figure 2e). Ordination of the plots from the two regions revealed that they span about the same range along PCoA axis 1 but are to some degree shifted in relation to each other along PCoA axis 2 (Figure 2f). The edaphic gradients covered by the plots were rather similar in both regions, and the relationship between the main floristic gradient (PCoA 1) and soil base cation concentration was even tighter than in the full dataset (compare Figure 2g with Figure 2c). The geographical distance between the Tigre and Juruá regions is related to differences in several of the CHELSA climatic variables. Compared to all of Amazonia, the climatic variation within and between the Tigre and Juruá areas is small, and the second PCoA axis was more weakly explained by the maximum temperature of the warmest month than in the case of the full dataset (compare Figure 2h with Figure 2d).

3.3 | Predicting species composition

The map based on the combined Landsat + CHELSA GLM suggests a general compositional difference between central–north-western Amazonia (blue–green in Figure 4a) and the peripheral areas (red–orange in Figure 4a). Much of this pattern is due to the strong spatial structure in predicted PCoA axis 1, which is assigned to red in Figure 4a and is highly correlated with soil base cation concentration. Indeed, the same pattern is even more clear in Figure 4b, which shows predictions for axis 1 using Landsat data only. Many of the known geological and vegetation characteristics already mentioned above (Pastaza fan, bamboo forests, Içá-Solimões boundary) were recovered in the PCoA axis 1 scores predicted by Landsat data (Figure 4b). Since these formations are not related to climatic differences, they became somewhat blurred when CHELSA data were included in the model (Figure 4a) even though the overall predictive power increased (Table 1).

A secondary general pattern can be identified as a NW–SE trend in core Amazonia. This emerged especially from variation in the predicted scores along PCoA axes 2 and 3, which in turn were mostly explainable by the climatic CHELSA data.

Zooming in to the regional scale highlights the inherent differences between the Landsat and CHELSA data. Patterns in the Landsat predictions can be easily related to local to regional landscape features that are also identifiable in the original Landsat image. For example, in the false-colour Landsat composite (Figure 5a) the swamp forests of the Pastaza fan are distinguishable as dark red-brown patches, and the limit between the forests growing on the

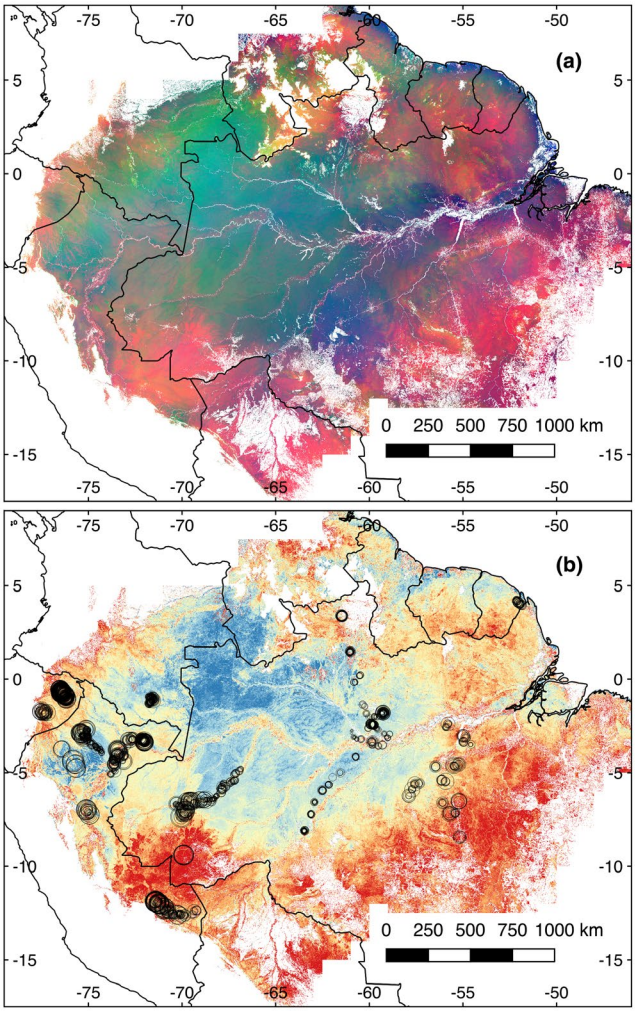

FIGURE 4 Colour composites over Amazonia showing GLM-predicted values for the PCoA axis scores of the fern and lycophyte ordination shown in Figure 2b. (a) Predictions based on both Landsat TM/ETM+ surface reflectance and CHELSA climate variables for the scores of PCoA axes 1, 2 and 3 (assigned to red, green and blue, respectively). (b) Predictions based on Landsat reflectance data only for the scores of PCoA axis 1 (shown in a colour gradient ranging from blue for low axis scores through yellow to red for high axis scores). The 1,572 field study plots used to parameterize the GLM models are shown as circles in (b), with diameter proportional to log-transformed base cation concentration in the soil. Pixels predicted to have extreme PCoA axis values (below −1.2 or above 1.2) were masked and appear in white (in addition to the areas masked already in Figure 1). GLM, generalized linear model; PCoA, Principal Coordinates Analysis

Nauta and Pebas Formations (dark green vs. lighter green with pinkish tint) is clearly visible. The same patterns appear as contrasting predictions for PCoA axis 1 scores when only Landsat data are used (Figure 5c). In contrast, climate varies much more gradually across space, so the patterns produced by climatic models have a weaker relationship with the local landscape (Figure 5b,d).

On the basis of the GLM-predicted floristic PCoA axes, we divided Amazonia into different numbers of geoecological classes using k-medoid clustering (CLARA). The number of classes that is chosen for use is of course arbitrary, but since large areas are entirely devoid of field data, we chose to focus on the broad-scale patterns and a relatively small number of classes. We found the solution with eight classes to provide a decent compromise between recognizing the most distinctive known macrounits and avoiding the creation of classes that cannot be substantiated with the data available to us (Figure 6). Further subdivision not only resulted in spatially very fragmented classes, but also produced boundaries that looked artifactual (e.g., by exactly following a line separating two temperature values in CHELSA). The resulting map is here presented as a hypothesis of a geoecological subdivision of Amazonia that can be further tested and refined with additional field and remote sensing data.

4 | DISCUSSION

4.1 | Determinants of plant species composition in Amazonia

In general, it is assumed that climate determines species distribution patterns at broad scales, and the effect of soils becomes noticeable at regional to local scales. However, we found that across entire Amazonia, the strongest floristic gradient in our field data corresponds to an edaphic gradient: a single soil variable (concentration of exchangeable base cations) explained as much as 75% of the variation in PCoA axis 1 values of ferns and lycophytes. It is likely that if other important soil properties could be included, such as phosphorus and nitrogen concentration or hydrology, the percentage of variation explained by soils would be even higher. This considerably expands the conclusions from earlier studies that have documented soils to be important for plant species turnover at regional extents in Amazonia (Baldeck et al., 2016; Cámara-Leret, Tuomisto, Ruokolainen, Balslev, & Munch Kristiansen, 2017; Higgins et al., 2011; Pansonato et al., 2013; Phillips et al., 2003; Ruokolainen et al., 2007; Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003; Tuomisto, Ruokolainen, & Yli-Halla, 2003; Zuquim et al., 2012, 2014).

Although soils explained the strongest floristic gradient in our data, climatic variables were also important. Their unique contribution was mostly in explaining the secondary floristic gradients (PCoA axes 2 and 3), for which soil base cation concentration provided no explanatory power at all. The relative importance of explanatory variables partly reflect the degree to which each of them varies in relation to the tolerances of the species of interest. Our entire study area is within the climatic space of moist tropical lowland forests and this may be a shorter gradient for plants than the observed edaphic one. In addition, soil base cation concentrations were measured from soil samples collected in the same plots as the floristic data. This may give a more accurate estimate of the conditions experienced by the plants than is the case with the CHELSA climatic variables, which have been derived from global climate models.

Because earlier studies have found strong edaphic relationships in many plant groups (ranging from canopy trees to understorey herbs), we suggest that soil variation needs to be taken into account

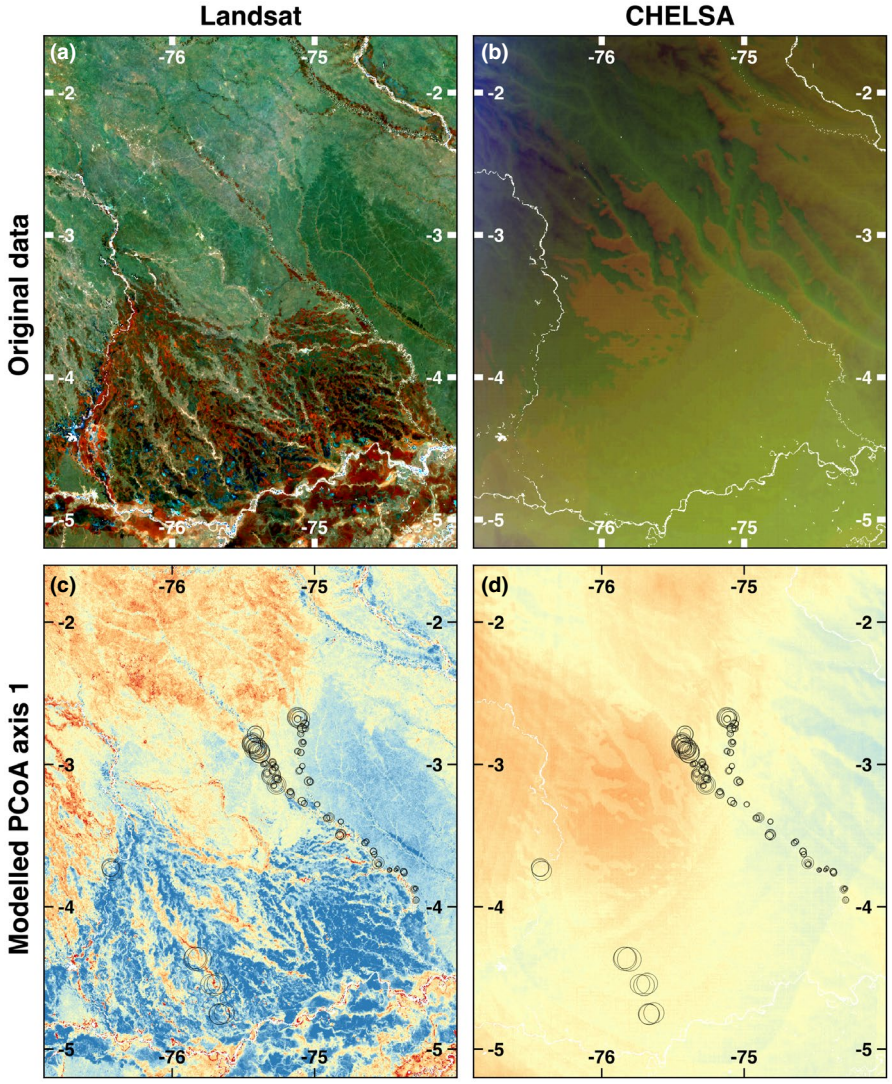

FIGURE 5 Landscape patterns in north-western Amazonia as modelled with either reflectance values from a Landsat TM/ETM+ image composite or bioclimatic variables from CHELSA. (a) Landsat colour composite based on bands 4, 5 and 7 assigned to red, green and blue, respectively. (b) CHELSA colour composite of three bioclimatic variables used to model PCoA axis 1 of the fern and lycophyte ordination shown in Figure 2b (Bio8, Bio11 and Bio2 assigned to red, green and blue, respectively). (c, d) Results of modelling the main floristic gradient of the area (PCoA axis 1 of Figure 2b) with Landsat data only (c) or CHELSA data only (d). The modelled values are shown in a colour gradient ranging from blue (for low scores along PCoA axis 1) through yellow to red (for high values). Each black circle represents one field plot with diameter proportional to log-transformed base cation concentration in the soil. Panel (a) is a detail of Figure S1b and panel (c) of Figure 4b. PCoA, Principal Coordinates Analysis

as a potentially important factor in all discussions about the distributional patterns of Amazonian plants. In particular, it can be expected that areas with contrasting soil properties have contrasting floristic compositions, and that the effects of dispersal limitation will accentuate if two areas with similar soils are separated by large expanses of different soils. These issues are especially important when considering how species distributions may be affected by deforestation or climate change: natural heterogeneity in soils may reduce the availability of suitable habitats even more than would be expected from deforestation or climate change scenarios alone (Figueiredo et al., 2018; Zuquim, Costa, Tuomisto, Moulatlet, & Figueiredo, 2019).

4.2 | Modelling species composition

The Landsat composite revealed clear geographical patterns, many of which obviously correspond to major geological formations. At the same time, the reflectance values provided reasonable predictions of both soil base cation concentration and the most important floristic gradient (PCoA axis 1) of understorey ferns and lycophytes. This is noteworthy, because surface reflectance over dense forests is mainly determined by the tree canopy, not by soils or understorey plants. Therefore, such a strong relationship is only possible if the floristic patterns in the understorey are causally linked with those vegetation properties that determine reflectance (including floristic composition, structure and chemical properties of the canopy). It seems clear that in our study area the link is mediated mostly by soils. Of course, without further studies we cannot establish to what degree the reflectance patterns mirror species-level differences in tree composition and to what degree similar structural or functional canopy properties on similar soils irrespective of species identity. Nevertheless, the causal chain from soils through vegetation to reflectance seems robust enough to make it possible to use Landsat data to identify ecologically relevant geological limits in densely vegetated areas.

The GLMs that used Landsat reflectance data predicted the main floristic gradient (fern and lycophyte PCoA Axis 1) better than they predicted soil base cation concentration. Landsat data also had some predictive power for PCoA axis 2, in contrast with soil base cation concentration, which had none. Both of these results are consistent with the idea that all plant groups react to both soils and climate in

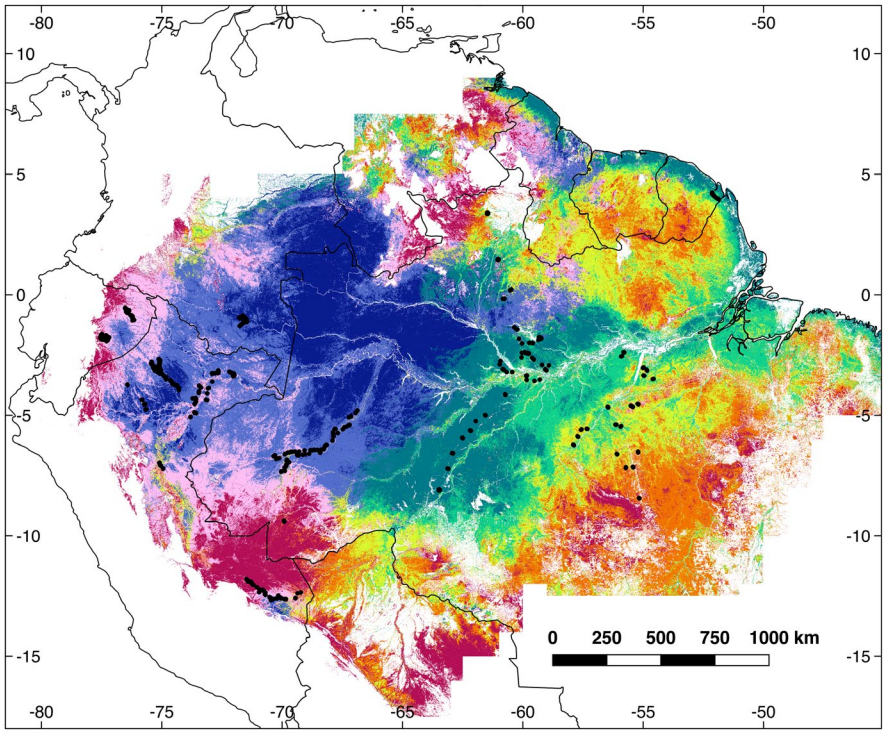

FIGURE 6 Classification of Amazonia into eight geoecological classes based on pixel-wise predictions of floristic ordination scores (PCoA axes 1–3 obtained with ferns and lycophytes; Figure 2b). Predictions were based on the same GLM as Figure 4a, which includes both Landsat reflectances and CHELSA bioclimatic variables as predictors. Black dots indicate the locations of the field plots that were used in the ordination. GLM, generalized linear model; PCoA, Principal Coordinates Analysis

ways that affect their floristic composition, structure and/or chemical properties, and that Landsat reflectance provides an integrated view of such environmental effects. Within Amazonia, Landsat reflectance mostly mirrors soil variation because the soil gradients are longer than the climatic gradients when compared to the physiological tolerances of Amazonian plants.

In two areas, there is an apparent discrepancy between the predictions of the GLMs and the actual position of the sites along PCoA axis 1 and the soil base cation concentration gradient. One is the Pastaza fan area, which is fed by rivers that originate in the Ecuadorian volcanoes and carry a high load of cation-rich sediments. All our soil samples from the fan have high base cation concentration (Figure 4b), but the area stands out as having low predicted PCoA axis 1 values, corresponding to forests on cation-poor soils. The Pastaza fan is very heterogeneous (Figure 5a,c), and our soil samples come from the narrow strips along the rivers that were, in fact, predicted to have high values. Soils further away from the rivers are probably both more cation-poor and more waterlogged, as the Pastaza fan is dominated by swamp forests. These may have lower reflectance both because infrared wavelengths are absorbed by water and because waterlogging makes swamps stressful environments, which can give them structural and chemical characteristics resembling those of forests on cation-poor soils. Some of the swamps have even evolved into ombrotrophic peat bogs, which are nutrient-limited in the same way as forests on white sand soils are, and have also been found to share structural characteristics and plant species (Draper et al., 2018; Lähteenoja & Page, 2011).

The second example is the opposite: the bamboo-dominated forests in the border zone between southern Peru and adjacent Brazil have high infrared reflectance, which is generally indicative of relatively rapid growth and forests on high-cation soils. Here, the estimates are probably exaggerated because of the bamboo in the canopy. Bamboo is a rapidly growing grass, and therefore can be expected to have less sclerophyllous leaves and higher infrared reflectance than canopy trees do. Although the fern and lycophyte plots we have from this area have cation-rich soils and a corresponding flora, the PCoA axis 1 predictions are more extreme than the floristic composition of the understorey would suggest.

4.3 | Basin-wide floristic mapping and the role of medium resolution multispectral imagery

Biodiversity studies covering all Amazonia have not used medium-resolution multispectral imagery, which is likely due to two main problems. Firstly, atmospheric contamination and persistent cloud cover hamper combining scenes acquired at different times. Secondly, directional scattering of sunlight by the canopy surface causes an artifactual along-scan (east-west) gradient in pixel values, which can cause spectral differences as large as those between compositionally different forest types (Muro et al., 2016; Toivonen, Kalliola, Ruokolainen, & Naseem Malik, 2006). However, recent advances in open data access policies (Woodcock et al., 2008), cloud screening and atmospheric correction algorithms (Masek et al., 2006), and directional normalization (Van doninck & Tuomisto, 2017a) now make it possible to construct seamless medium resolution image composites with a reasonably high signal-to-noise ratio over large areas (Van doninck & Tuomisto, 2018).

In the Landsat composite (Figure 1), reflectance variation mostly corresponds to identifiable surface features. Our modelling results (Table 1) confirmed that multispectral surface reflectance data

layers make a useful contribution to modelling understorey fern and lycophyte species composition at the basin-wide extent, complementing information that can be obtained from climate variables. This is in accordance with results of several studies carried out at relatively small extents (Buermann et al., 2008; Chaves, Ruokolainen, & Tuomisto, 2018; Figueiredo, Venticinque, Figueiredo, & Ferreira, 2015; Higgins et al., 2012, 2011; Salovaara et al., 2005; Thessler et al., 2005; Tuomisto, Poulsen, et al., 2003; Tuomisto, Ruokolainen, Aguilar, et al., 2003). New generation medium-resolution multispectral instruments with improved spectral and radiometric resolution (Landsat 8 OLI, Sentinel-2) can be expected to be even more valuable for these applications. Consequently, we disagree with the suggestion that the spatial and spectral resolution of this type of sensors would be inappropriate for studies on spatial distribution of biodiversity or trait variations (Lausch et al., 2016; Nagendra & Rocchini, 2008).

The choice of spatial resolution in remote sensing studies determines what information on biodiversity can be gained (Anderson, 2018; Rocchini et al., 2016). Our results show that when the aim is to map broad-scale patterns in floristic composition, valuable information can be extracted without high-resolution imagery. Even at medium resolution, a single image pixel is about the size of a large tree crown or a treefall gap in a tropical forest. This causes a high degree of local variability, because adjacent pixels can represent different phases of gap dynamics in different proportions, which makes the identification of general patterns more difficult. Filtering is the classical method for eliminating local noise and it has been used in earlier studies in Amazonia (Chaves et al., 2018; Salovaara et al., 2005). We indeed found that applying a 15-pixel median filter considerably improved model performance, indicating that when the aim is broad-scale floristic mapping, the high heterogeneity between adjacent 30-m pixels is mostly noise.

Nevertheless, it is an advantage to have access to the medium-resolution data and not only coarse-resolution data. Firstly, this allows spectral values to be extracted such that they are centred on the exact locality of field sampling points, rather than the field data being potentially marginally placed in relation to a large pre-defined pixel. Secondly, there may be non-forest land cover types such as roads, rivers or cultivated fields close to the field sampling locations. With coarse-resolution imagery, these would lead to mixed pixels, but with medium resolution imagery, one can mask out the irrelevant pixels before extracting the reflectance values. Finally, medium resolution data allows generating entropy or variability metrics, which may be indicative of local taxonomic diversity (Rocchini et al., 2018), even though here we found the simple standard deviation metric to be uninformative.

4.4 | Biogeographical inferences and practical applications

Our results show that Landsat reflectance can be used to generalize field data and to predict soil-related floristic variation at the basin-wide extent. Thereby, Landsat provides information that goes beyond and complements climatic data. Although our maps are based on modelling fern and lycophyte species compositional gradients, patterns identified by Landsat are hardly specific to these understorey plants. Therefore, we expect the information in our maps to be relevant for Amazonian biota more generally, including other plant groups and those animal groups that react to spatial variation in soil-related forest properties. Further work is needed to test to what degree the patterns identified here apply for other taxa and in areas for which we had no field data. Nevertheless, this mapping approach is based on solid ecological principles and opens exciting possibilities for future ecological and biogeographical research with implications for how we view Amazonia and the threats it is facing.

In general terms, the classical division of Amazonia into four geochemically defined regions as proposed by Fittkau et al. (1975) is discernible in our results. In the geoecological classification of Figure 6, the red, pink and pale blue classes roughly correspond to western Amazonia, dark blue and green classes to central Amazonia, and orange and yellow classes to southern Amazonia and the Guiana Shield. However, the southern part of the western periphery appears both in the Landsat composite (Figure 1) and in the geoecological classification (Figure 6) as more related to southern Amazonia than to the northern parts of western Amazonia, thus supporting the subdivision of the latter as proposed by ter Steege et al. (2013).

The western limit of central Amazonia in our results coincides with a previously identified geological and floristic boundary (Higgins et al., 2011; IBGE, 2004; Schobbenhaus et al., 2004; Tuomisto et al., 2016). Central Amazonia as defined by ter Steege et al. (2013) extends further west than this, whereas in the north their boundary is south of the extensive white sand areas (Adeney et al., 2016; Quesada et al., 2011) that in both the Landsat composite and the geoecological classification associate with central Amazonia. The eastern Amazonian region recognized by ter Steege et al. (2013) around the mouth of the Amazon river stands out also in our Landsat composite (Figure 1), and was allocated to a different class than most of southern Amazonia (Figure 6). It must be noted, however, that lack of field data renders our results for this area (and for southern Amazonia in general) rather speculative.

The limits between geoecological subregions mostly did not follow major rivers, with the exception of the upper Rio Negro and, to some degree, the lower Amazon itself (Figures 1 and 6). Traditionally, the Amazon River and its main tributaries have been recognized as distributional limits for many animal species, and it has been suggested that the rivers function as dispersal barriers (Aleixo, 2006; Godinho & da Silva, 2018; Haffer, 1974; Nazareno, Dick, & Lohmann, 2017; Pomara, Ruokolainen, & Young, 2014; Ribas et al., 2012; Wallace, 1852). Where a river coincides with habitat differences, it is difficult to disentangle the possible effects of a river barrier from those of habitat selection, but in areas where rivers and habitat limits do not coincide, the two hypotheses make different predictions depending on species vagility and degree of habitat specificity. This has an important practical consequence: any regional maps where category limits are drawn using rivers as boundaries (such as the

WWF ecoregion map; Olson et al., 2001) may only be relevant for river-limited organisms.

A few recent studies have used Landsat data as environmental layers in species distribution models (SDMs) over relatively small extents (Chaves et al., 2018; Figueiredo et al., 2015). Our results suggest that the same approach is feasible even at the basin-wide extent (see also Van doninck & Tuomisto, 2018). For example, the modelled floristic gradient shown in Figure 4a allows making rather specific predictions about the potential distributions of plant species: only species tolerant of cation-poor soils are expected to grow in the blue-green areas in central Amazonia, while increasing redness in the map indicates increasing probability of occurrence for species requiring high cation availability. Until now, SDM has often been done using climatic variables only, and the few studies that have used digital soil maps (Figueiredo et al., 2018; Levis et al., 2017) may have underestimated the importance of soils due to the problems with thematic and spatial accuracy in the available soil maps (Moulatlet et al., 2017).

Because our models focus on the dominant floristic gradients only, and are based on plants with relatively good dispersal ability, they have been trained to emphasize environmental site conditions. The degree of floristic regionalization that emerges through isolation by distance is probably underestimated even for ferns and lycophytes, and more so for plant groups that are more dispersal-limited. This will need to be taken into account when assessing the ecological and biogeographical significance of the subdivisions of Amazonia that emerge from our results (especially Figure 6). Our results provide one geoecological view over Amazonia, and comparable studies using other plant groups are now needed to test this view.

ACKNOWLEDGEMENTS

We are grateful to the numerous people who have contributed to the data used in this paper by participating in fieldwork, helping with practical arrangements for the field expeditions or sharing their expertise for species identification or soil analysis. We thank the national authorities in each country for granting the permits to carry out fieldwork and collect voucher specimens. We thank Hans ter Steege, Ethan Householder and Dilce Rossetti for constructive comments that helped improve the manuscript. Funding that has made this work possible has been provided by numerous funding agencies over the years, most recently by the Academy of Finland (grants 139959 and 273737 to H.T., grant 296406 to Risto Kalliola), Finnish Cultural Foundation (grant to G.Z.) and several grants associated with the Brazilian Program in Biodiversity Research (PPBio) from Brazilian agencies like FAPEAM and CNPq. This is a publication of the Amazon Research Team of the University of Turku, and also contributes as publication 759 to the technical series of the Biological Dynamics of Forest Fragments Project.

DATA AVAILABILITY STATEMENT

The plot data used in this paper have been deposited in the Dryad Digital Repository at: https://doi.org/10.5061/dryad.v7fp8ms and will be released in 2026. The Landsat TM/ETM+ composite has been deposited in IDA Research data storage service (www.fairdata.fi/en/ida/). Researchers interested in the data should contact Hanna Tuomisto (for either dataset) or Jasper Van doninck (for the Landsat dataset).

Title: Data from: Discovering floristic and geoecological gradients across AmazoniaDOI: doi:10.5061/dryad.v7fp8ms

Journal: Journal of BiogeographyJournal manuscript number: JBI-18-0631

ORCID

Hanna Tuomisto https://orcid.org/0000-0003-1640-490X
Jasper Van doninck https://orcid.org/0000-0003-2177-7882
Kalle Ruokolainen https://orcid.org/0000-0002-7494-9417
Gabriel M. Moulatlet https://orcid.org/0000-0003-2571-1207
Fernando O. G. Figueiredo https://orcid.org/0000-0002-9333-0708
Anders Sirén https://orcid.org/0000-0003-4159-4506
Glenda Cárdenas https://orcid.org/0000-0002-3441-4602
Samuli Lehtonen https://orcid.org/0000-0001-6235-2026
Gabriela Zuquim https://orcid.org/0000-0003-0932-2308

REFERENCES

Adeney, J. M., Christensen, N. L., Vicentini, A., & Cohn-Haft, M. (2016). White-sand ecosystems in Amazonia. *Biotropica*, *48*(1), 7–23. https://doi.org/10.1111/btp.12293

Aleixo, A. (2006). Historical diversification of floodplain forest specialist species in the Amazon: A case study with two species of the avian genus Xiphorhynchus (Aves: Dendrocolaptidae). *Biological Journal of the Linnean Society*, *89*(2), 383–395. https://doi.org/10.1111/j.1095-8312.2006.00703.x

Anderson, C. B. (2018). Biodiversity monitoring, earth observations and the ecology of scale. *Ecology Letters*, *21*(10), 1572–1585. https://doi.org/10.1111/ele.13106

Asner, G. P., Anderson, C. B., Martin, R. E., Tupayachi, R., Knapp, D. E., & Sinca, F. (2015). Landscape biogeochemistry reflected in shifting distributions of chemical traits in the Amazon forest canopy. *Nature Geoscience*, *8*(7), 567–573. https://doi.org/10.1038/ngeo2443

Asner, G. P., Kellner, J. R., Kennedy-Bowdoin, T., Knapp, D. E., Anderson, C., & Martin, R. E. (2013). Forest canopy gap distributions in the southern Peruvian Amazon. *PLoS ONE*, *8*(4), e60875. https://doi.org/10.1371/journal.pone.0060875

Asner, G. P., Knapp, D. E., Martin, R. E., Tupayachi, R., Anderson, C. B., Mascaro, J., … Silman, M. R. (2014). Targeted carbon conservation at national scales with high-resolution monitoring. *Proceedings of the National Academy of Sciences*, *111*(47), E5016–E5022. https://doi.org/10.1073/pnas.1419550111

Baldeck, C. A., Tupayachi, R., Sinca, F., Jaramillo, N., & Asner, G. P. (2016). Environmental drivers of tree community turnover in western Amazonian forests. *Ecography*, *39*(11), 1089–1099. https://doi.org/10.1111/ecog.01575

Buermann, W., Saatchi, S., Smith, T. B., Zutta, B. R., Chaves, J. A., Milá, B., & Graham, C. H. (2008). Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, *35*(7), 1160–1176. https://doi.org/10.1111/j.1365-2699.2007.01858.x

Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H., & Munch Kristiansen, S. (2017). Modelling responses of western Amazonian palms to soil nutrients. *Journal of Ecology*, *105*(2), 367–381. https://doi.org/10.1111/1365-2745.12708

de Carvalho, A. L., Nelson, B. W., Bianchini, M. C., Plagnol, D., Kuplich, T. M., & Daly, D. C. (2013). Bamboo-dominated forests of the southwest Amazon: Detection, spatial extent, life cycle length and flowering waves. *PLoS ONE*, *8*(1), e54852. https://doi.org/10.1371/journal.pone.0054852

Chaves, P. P., Ruokolainen, K., & Tuomisto, H. (2018). Using remote sensing to model tree species distribution in Peruvian lowland Amazonia. *Biotropica*, *50*(5), 758–767. https://doi.org/10.1111/btp.12597

Cracraft, J. (1985). Historical biogeography and patterns of differentiation within the South American avifauna: Areas of endemism. *Ornithological Monographs*, *36*, 49–84. https://doi.org/10.2307/40168278

De'ath, G. (1999). Extended dissimilarity: A method of robust estimation of ecological distances from high beta diversity data. *Plant Ecology*, *144*(2), 191–199. https://doi.org/10.1023/A:1009763730207

Draper, F. C., Baraloto, C., Brodrick, P. G., Phillips, O. L., Martinez, R. V., Honorio Coronado, E. N., … Asner, G. P. (2019). Imaging spectroscopy predicts variable distance decay across contrasting Amazonian tree communities. *Journal of Ecology*, *107*(2), 696–710. https://doi.org/10.1111/1365-2745.13067

Draper, F. C., Honorio Coronado, E. N., Roucoux, K. H., Lawson, I. T., A. Pitman, N. C., A. Fine, P. V., … Baker, T. R. (2018). Peatland forests are the least diverse tree communities documented in Amazonia, but contribute to high regional beta-diversity. *Ecography*, *41*(8), 1256–1269. https://doi.org/10.1111/ecog.03126

Duivenvoorden, J. F., & Lips, J. M. (1993). *Ecología del paisaje del Medio Caquetá. Memoria explicativa de los mapas*. Santa Fé de Bogotá, Colombia: Tropenbos Colombia.

Duivenvoorden, J. F. (1995). Tree species composition and rain forest-environment relationships in the middle Caquetá area, Colombia, NW Amazonia. *Vegetatio*, *120*(2), 91–113.

Duque, A. J., Duivenvoorden, J. F., Cavelier, J., Sánchez, M., Polanía, C., & León, A. (2005). Ferns and Melastomataceae as indicators of vascular plant composition in rain forests of Colombian Amazonia. *Plant Ecology*, *178*(1), 1–13. https://doi.org/10.1007/s11258-004-1956-2

Figueiredo, F. O. G., Zuquim, G., Tuomisto, H., Moulatlet, G. M., Balslev, H., & Costa, F. R. C. (2018). Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species. *Journal of Biogeography*, *45*(1), 190–200. https://doi.org/10.1111/jbi.13104

Figueiredo, S. M. de M., Venticinque, E. M., Figueiredo, E. O., & Ferreira, E. J. L. (2015). Predicting the distribution of forest tree species using topographic variables and vegetation index in eastern Acre, Brazil. *Acta Amazonica*, *45*(2), 167–174. https://doi.org/10.1590/1809-4392201402834

Fittkau, E. J., Junk, J. W., Klinge, H., & Sioli, H. (1975). Substrate and vegetation in the Amazon region. In H. Dierschke, & R. Tüxen (Eds.), *Vegetation und Substrat* (pp. 73–90). Vaduz (Liechtenstein): J. Cramer.

Godinho, M. B. de C., & da Silva, F. R. (2018). The influence of riverine barriers, climate, and topography on the biogeographic regionalization of Amazonian anurans. *Scientific Reports*, *8*(1), 3427. https://doi.org/10.1038/s41598-018-21879-9

Haffer, J. (1974). *Avian speciation in tropical South America, with a systematic survey of the toucans (Ramphastidae) and jacamars (Galbulidae)* (1st ed.). Cambridge, Mass: Nuttall Ornithological Club.

Halme, K. J., & Bodmer, R. E. (2007). Correspondence between scientific and traditional ecological knowledge: Rain forest classification by the non-indigenous ribereños in Peruvian Amazonia. *Biodiversity and Conservation*, *16*(6), 1785–1801. https://doi.org/10.1007/s10531-006-9071-4

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., … Gonzalez, M. R. (2014). SoilGrids1km – Global soil information based on automated mapping. *PLoS ONE*, *9*(8), e105992. https://doi.org/10.1371/journal.pone.0105992

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Higgins, M. A., Asner, G. P., Perez, E., Elespuru, N., Tuomisto, H., Ruokolainen, K., & Alonso, A. (2012). Use of Landsat and SRTM data to detect broad-scale biodiversity patterns in northwestern Amazonia. *Remote Sensing*, *4*(12), 2401–2418. https://doi.org/10.3390/rs4082401

Higgins, M. A., 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*(11), 2136–2149. https://doi.org/10.1111/j.1365-2699.2011.02585.x

Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling (Version 2.8-19). Retrieved from https://CRAN.R-project.org/package=raster

Hoorn, C. (1993). Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: Results of a palynostratigraphic study. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *105*(3), 267–309. https://doi.org/10.1016/0031-0182(93)90087-Y

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., … Antonelli, A. (2010). Amazonia through time: andean uplift, climate change, landscape evolution, and biodiversity. *Science*, *330*(6006), 927–931. https://doi.org/10.1126/science.1194585

Huber, O., & Alarcón, C. (1988). *Mapa de vegetación de Venezuela*. Caracas, Venezuela: Ministerio del Ambiente y de los Recursos Naturales Renovables.

Huber, O., Gharbarran, G., & Funk, V. (1995). *Vegetation map of Guyana (preliminary version)*. Retrieved from http://botany.si.edu/bdg/vegmap.html

IBGE. (2004). *Mapa de Vegetação do Brasil 1:5,000,000* (3rd ed.). Retrieved from fttp://geoftp.ibge.gov.br/mapas/tematicos/mapas_murais

Jones, M. M., Ferrier, S., Condit, R., Manion, G., Aguilar, S., & Pérez, R. (2013). Strong congruence in tree and fern community turnover in response to soils and climate in central Panama. *Journal of Ecology*, *101*(2), 506–516. https://doi.org/10.1111/1365-2745.12053

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., … Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122. https://doi.org/10.1038/sdata.2017.122

Lähteenoja, O., & Page, S. (2011). High diversity of tropical peatland ecosystem types in the Pastaza-Marañón basin, Peruvian Amazonia. *Journal of Geophysical Research: Biogeosciences*, *116*(G2), G02025. https://doi.org/10.1029/2010JG001508

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J. M., … Cord, A. F. (2016). Linking earth observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators*, *70*, 317–339. https://doi.org/10.1016/j.ecolind.2016.06.022

Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., … ter Steege, H. (2017). Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science*, *355*(6328), 925–931. https://doi.org/10.1126/science.aal0157

Magnusson, W. E., Lima, A. P., Luizão, R., Luizão, F., Costa, F. R. C., de Castilho, C. V., & Kinupp, V. F. (2005). RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. *Biota Neotropica*, *5*, 19–24.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2016). cluster: Cluster analysis basics and extensions (Version R package version 2.0.4). Retrieved from https://cran.r-project.org/web/packages/cluster/citation.html

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., … Lim, T.-K. (2006). A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68–72. https://doi.org/10.1109/LGRS.2005.857030

Mitchard, E. T. A., Feldpausch, T. R., Brienen, R. J. W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., … Phillips, O. L. (2014). Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global Ecology and Biogeography*, *23*(8), 935–946. https://doi.org/10.1111/geb.12168

Moulatlet, G. M., Zuquim, G., Figueiredo, F. O. G., Lehtonen, S., Emilio, T., Ruokolainen, K., & Tuomisto, H. (2017). Using digital soil maps to infer edaphic affinities of plant species in Amazonia: Problems and prospects. *Ecology and Evolution*, *7*(20), 8463–8477. https://doi.org/10.1002/ece3.3242

Muro, J., Van doninck, J., Tuomisto, H., Higgins, M. A., Moulatlet, G. M., & Ruokolainen, K. (2016). Floristic composition and across-track reflectance gradient in Landsat images over Amazonian forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 361–372. https://doi.org/10.1016/j.isprsjprs.2016.06.016

Nagendra, H., & Rocchini, D. (2008). High resolution satellite imagery for tropical biodiversity studies: The devil is in the detail. *Biodiversity and Conservation*, *17*(14), 3431. https://doi.org/10.1007/s10531-008-9479-0

Nazareno, A. G., Dick, C. W., & Lohmann, L. G. (2017). Wide but not impermeable: Testing the riverine barrier hypothesis for an Amazonian plant species. *Molecular Ecology*, *26*(14), 3636–3648. https://doi.org/10.1111/mec.14142

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., …Wagner, H. (2017). vegan: Community ecology package (Version 2.4-5). Retrieved from https://CRAN.R-project.org/package=vegan

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*(11), 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Pansonato, M. P., Costa, F. R. C., de Castilho, C. V., Carvalho, F. A., & Zuquim, G. (2013). Spatial scale or amplitude of predictors as determinants of the relative importance of environmental factors to plant community structure. *Biotropica*, *45*(3), 299–307. https://doi.org/10.1111/btp.12008

Peres, C. A. (1999). Effects of subsistence hunting and forest types on the structure of Amazonian primate communities. In J. G. Fleagle, C. Janson, & K. E. Reed (Ed.), *Primate communities* (pp. 268–283). Cambridge, UK: Cambridge University Press.

Phillips, O. L., Vargas, P. N., Monteagudo, A. L., Cruz, A. P., Zans, M.-E., Sánchez, W. G., … Rose, S. (2003). Habitat association among Amazonian tree species: A landscape-scale approach. *Journal of Ecology*, *91*(5), 757–775. https://doi.org/10.1046/j.1365-2745.2003.00815.x

Pomara, L. Y., Ruokolainen, K., & Young, K. R. (2014). Avian species composition across the Amazon River: The roles of dispersal limitation and environmental heterogeneity. *Journal of Biogeography*, *41*(4), 784–796. https://doi.org/10.1111/jbi.12247

Quesada, C. A., Lloyd, J., Anderson, L. O., Fyllas, N. M., Schwarz, M., & Czimczik, C. I. (2011). Soils of Amazonia with particular reference to the RAINFOR sites. *Biogeosciences*, *8*(6), 1415–1440. https://doi.org/10.5194/bg-8-1415-2011

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Räsänen, M. E., Linna, A. M., Santos, J. C. R., & Negri, F. R. (1995). Late Miocene tidal deposits in the Amazonian foreland basin. *Science*, *269*(5222), 386–390. https://doi.org/10.1126/science.269.5222.386

Räsänen, M., Neller, R., Salo, J., & Jungner, H. (1992). Recent and ancient fluvial deposition systems in the Amazonian foreland basin, Peru. *Geological Magazine*, *129*(03), 293–306. https://doi.org/10.1017/S0016756800019233

Ribas, C. C., Aleixo, A., Nogueira, A. C. R., Miyaki, C. Y., & Cracraft, J. (2012). A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 681–689. https://doi.org/10.1098/rspb.2011.1120

Rocchini, D., Boyd, D. S., Féret, J.-B., Foody, G. M., He, K. S., Lausch, A., … Pettorelli, N. (2016). Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sensing in Ecology and Conservation*, *2*(1), 25–36. https://doi.org/10.1002/rse2.9

Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., … Nagendra, H. (2018). Measuring β-diversity by remote sensing: A challenge for biodiversity monitoring. *Methods in Ecology and Evolution*, *9*(8), 1787–1798. https://doi.org/10.1111/2041-210X.12941

Rossetti, D. D. F., Cassola Molina, E., & Cremon, É. H. (2016). Genesis of the largest Amazonian wetland in northern Brazil inferred by morphology and gravity anomalies. *Journal of South American Earth Sciences*, *69*, 1–10. https://doi.org/10.1016/j.jsames.2016.03.006

Ruokolainen, K., Tuomisto, H., Macía, M. J., Higgins, M. A., & Yli-Halla, M. (2007). Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology*, *23*(01), 13–25. https://doi.org/10.1017/S0266467406003889

Saatchi, S., Malhi, Y., Zutta, B., Buermann, W., Anderson, L. O., Araujo, A. M., … Ramírez, H. A. (2009). Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia. *Biogeosciences Discuss*. *6*(3), 5461–5505. https://doi.org/10.5194/bgd-6-5461-2009

Saatchi, S. S., Houghton, R. A., Dos Santos Alvalá, R. C., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, *13*(4), 816–837. https://doi.org/10.1111/j.1365-2486.2007.01323.x

Salovaara, K. J., Thessler, S., Malik, R. N., & Tuomisto, H. (2005). Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment*, *97*(1), 39–51. https://doi.org/10.1016/j.rse.2005.04.013

Sanchez, P. A., & Buol, S. W. (1974). Properties of some soils of the upper Amazon basin of Peru. *Soil Science Society of America Journal*, *38*(1), 117. https://doi.org/10.2136/sssaj1974.03615995003800010036x

Schobbenhaus, C., Gonçalves, J. H., Santos, J. O. S., Abram, M. B., Leao Neto, R., Matos, G. M. M., …Jesus, J. D. A. (2004). Carta geológica do Brasil ao milionésimo, sistemas de informações geográficas-SIG/Geological map of Brazil, 1:1,000,000 scale, geographic information system-GIS. CD-ROM. CPRM, Geological Survey of Brazil, Brasilia.

Sirén, A., Tuomisto, H., & Navarrete, H. (2013). Mapping environmental variation in lowland Amazonian rainforests using remote sensing and floristic data. *International Journal of Remote Sensing*, *34*(5), 1561–1575. https://doi.org/10.1080/01431161.2012.723148

Sombroek, W. (2000). Amazon landforms and soils in relation to biological diversity. *Acta Amazonica*, *30*(1), 81–100. https://doi.org/10.1590/1809-43922000301100

Stevenson, P. R. (2014). Potential determinants of the abundance of woolly monkeys in neotropical forests. In Defler, T. R., & Stevenson, P. R. (Ed.). *The Woolly Monkey* (pp. 207–226). New York, NY: Springer.

Stropp, J., ter Steege, H., Malhi, Y., ATDN, & RAINFOR. (2009). Disentangling regional and local tree diversity in the Amazon. *Ecography*, *32*(1), 46–54. https://doi.org/10.1111/j.1600-0587.2009.05811.x

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*(7110), 444–447. https://doi.org/10.1038/nature05134

ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., … Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science*, *342*(6156), 1243092. https://doi.org/10.1126/science.1243092

ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., van der Hout, P., Daly, D. C., … Morawetz, W. (2003). A spatial model of tree α-diversity and tree density for the Amazon. *Biodiversity & Conservation*, *12*(11), 2255–2277. https://doi.org/10.1023/A:1024593414624

Thessler, S., Ruokolainen, K., Tuomisto, H., & Tomppo, E. (2005). Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. *Global Ecology and Biogeography*, *14*(4), 315–325. https://doi.org/10.1111/j.1466-822X.2005.00158.x

Thiers, B. (continuously updated). Index Herbariorum: A global directory of public herbaria and associated staff. Retrieved from New York Botanical Garden's Virtual Herbarium. Website: http://sweetgum.nybg.org/ih/

Toivonen, T., Kalliola, R., Ruokolainen, K., & Naseem Malik, R. (2006). Across-path DN gradient in Landsat TM imagery of Amazonian forests: A challenge for image interpretation and mosaicking. *Remote Sensing of Environment*, *100*(4), 550–562. https://doi.org/10.1016/j.rse.2005.11.006

Tuomisto, H., Moulatlet, G. M., 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*, *43*(12), 2400–2411. https://doi.org/10.1111/jbi.12864

Tuomisto, H., Poulsen, A. D., Ruokolainen, K., Moran, R. C., Quintana, C., Celi, J., & Cañas, G. (2003). Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. *Ecological Applications*, *13*(2), 352–371. https://doi.org/10.1890/1051-0761(2003) 013[0352:LFPWSH]2.0.CO;2

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*(5), 743–756. https://doi.org/10.1046/j.1365-2745.2003.00802.x

Tuomisto, H., Ruokolainen, K., Vormisto, J., Duque, A., Sánchez, M., Vargas Paredes, V., & Lähteenoja, O. (2017). Effect of sampling grain on patterns of species richness and turnover in Amazonian forests. *Ecography*, *40*(7), 840–852. https://doi.org/10.1111/ecog.02453

Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, *299*(5604), 241–244. https://doi.org/10.1126/science.1078037

Tuomisto, H., Ruokolainen, L., & Ruokolainen, K. (2012). Modelling niche and neutral dynamics: On the ecological interpretation of variation partitioning results. *Ecography*, *35*(11), 961–971. https://doi.org/10.1111/j.1600-0587.2012.07339.x

Van doninck, J., & Tuomisto, H. (2017a). Evaluation of directional normalization methods for Landsat TM/ETM+ over primary Amazonian lowland forests. *International Journal of Applied Earth Observation and Geoinformation*, *58*, 249–263. https://doi.org/10.1016/j.jag.2017.01.017

Van doninck, J., & Tuomisto, H. (2017b). Influence of compositing criterion and data availability on pixel-based Landsat TM/ETM+ image compositing over Amazonian forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(3), 857–867. https://doi.org/10.1109/JSTARS.2016.2619695

Van doninck, J., & Tuomisto, H. (2018). A Landsat composite covering all Amazonia for applications in ecology and conservation. *Remote Sensing in Ecology and Conservation*, *4*(3), 197–210. https://doi.org/10.1002/rse2.77

Vormisto, J., Phillips, O. L., Ruokolainen, K., Tuomisto, H., & Vásquez, R. (2000). A comparison of fine-scale distribution patterns of four plant groups in an Amazonian rainforest. *Ecography*, *23*(3), 349–359. https://doi.org/10.1111/j.1600-0587.2000.tb00291.x

Wagner, F. H., Hérault, B., Rossi, V., Hilker, T., Maeda, E. E., Sanchez, A., … Aragão, L. E. O. C. (2017). Climate drivers of the Amazon forest greening. *PLoS ONE*, *12*(7), e0180932. https://doi.org/10.1371/journal.pone.0180932

Wallace, A. R. (1852). On the monkeys of the Amazon. *Proceedings of the Zoological Society of London*, *20*, 107–110.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., … Wynne, R. (2008). Free access to Landsat imagery. *Science*, *320*(5879), 1011a. https://doi.org/10.1126/science.320.5879.1011a

Zuquim, G., Costa, F. R. C., Tuomisto, H., Moulatlet, G. M., & Figueiredo, F. O. G. (2019). The importance of soils in predicting the future of plant habitat suitability in a tropical forest. *Plant and Soil*, 00, 1–20. https://doi.org/10.1007/s11104-018-03915-9

Zuquim, G., Tuomisto, H., Costa, F. R. C., Prado, J., Magnusson, W. E., Pimentel, T., … Figueiredo, F. O. G. (2012). Broad Scale distribution of ferns and lycophytes along environmental gradients in central and Northern Amazonia, Brazil. *Biotropica*, *44*(6), 752–762. https://doi.org/10.1111/j.1744-7429.2012.00880.x

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*(5), 1195–1207. https://doi.org/10.1111/jvs.12174

BIOSKETCH

Hanna Tuomisto, **Kalle Ruokolainen**, **Gabriela Zuquim** and most of the other authors are biologists interested in how species are distributed in Amazonia, what the determinants of species occurrences are and how these relate to the broader context of geological history and evolutionary processes. **Jasper Van doninck** is a geographer specialized on using remote sensing for biodiversity mapping.

Author contributions: H.T. and K.R. conceived the original idea, which was then refined by H.T, J.V.D., G.Z., K.R. and A.S.; H.T., K.R., G.Z., G.M.M., F.O.G.F., A.S., G.C. and S.L. collected the field data; J.V.D. and H.T. carried out the analyses; H.T., G.Z. and J.V.D. led the writing and all authors commented on the manuscript.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Tuomisto H, Van doninck J, Ruokolainen K, et al. Discovering floristic and geoecological gradients across Amazonia. *J Biogeogr.* 2019;00:1–15. https://doi.org/10.1111/jbi.13627