



OPEN ACCESS

EDITED BY

Debora Zurro,
Spanish National Research Council
(CSIC), Spain

REVIEWED BY

Aissam Bekkari,
Cadi Ayyad University, Morocco
Andre Carlo Colonese,
Autonomous University of Barcelona, Spain

*CORRESPONDENCE

Barney Harris
✉ bharris@bournemouth.ac.uk

RECEIVED 04 March 2025

ACCEPTED 26 August 2025

PUBLISHED 20 October 2025

CITATION

Harris B, Behling H, Gregorio de Souza J,
Reinhardt AL, Roberts P and Riris P (2025)
Investigating pre-Columbian floristic legacy
effects using machine learning in the southern
Atlantic Forest, Brazil.
Front. Environ. Archaeol. 4:1587670.
doi: 10.3389/fearc.2025.1587670

COPYRIGHT

© 2025 Harris, Behling, Gregorio de Souza,
Reinhardt, Roberts and Riris. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Investigating pre-Columbian floristic legacy effects using machine learning in the southern Atlantic Forest, Brazil

Barney Harris ^{1*}, Hermann Behling ²,
Jonas Gregorio de Souza ³, Antonia Lena Reinhardt ²,
Patrick Roberts ⁴ and Philip Riris ¹

¹Department of Archaeology & Anthropology, Bournemouth University, Poole, United Kingdom,

²Department of Palynology and Climate Dynamics, University of Goettingen, Göttingen, Germany,

³CASEs, Department of Humanities, Universitat Pompeu Fabra, Barcelona, Spain, ⁴Department of
Coevolution of Land Use and Urbanisation, Max Planck Institute of Geoanthropology, Jena, Germany

The scope and scale of past human impacts on both historic and current vegetation is of widespread interest in the historical sciences. In the Atlantic Forest of southern Brazil (Portuguese: Mata Atlântica), previous work has identified Amerindian settlement and land-use as a probable driver of the extent and composition of forest cover, with time-extended legacies that remain detectable in modern floristic inventories. Previously published investigations into the ecological history of the southern Atlantic Forest have either eschewed the role of humans or, where anthropogenic drivers are explicitly examined, utilized spatially restricted environmental datasets, necessarily limiting the generalizability their conclusions. This study aims to redress this gap, and to quantify the impact of past Amerindian Pre-Columbian settlement and associated land use on the modern-day distribution of several key plant species across the entire southern Atlantic Forest. We fit Maxent species distribution models (SDMs) using Indigenous archaeological site locations (Tupi-Guarani and southern Jê) and modern plant species occurrence data (35 unique species) in a comparative analytical framework to investigate Indigenous influence on the likelihood of occurrence of culturally significant or medicinal plant species. Our results indicate that (i) the inclusion of archaeological settlement location data and SDM predictions as covariates can improve the performance of contemporary floristic species distribution modeling and should be incorporated into ecological models of plant species in landscapes with long-standing human presence, especially when they are used to inform policy that explicitly aims to preserve “natural” biomes and; (ii) a synanthropic relationship can be demonstrated between the southern Jê and *Araucaria angustifolia*, a finding that complements previously published phylogeographic and palaeoenvironmental studies exploring the same link.

KEYWORDS

SDM, Atlantic Forest, geospatial, archaeoecology, Maxent

1 Introduction

The historic role of humans in shaping tropical forests, and their consequences for land use, land cover, and Earth system change, is a subject of significant current interest across the social and environmental sciences (Barlow et al., 2012; Malhi et al., 2014; Roberts P. et al., 2017; Roberts et al., 2023, 2018). In the Neotropics, emerging evidence has highlighted the degree to which, and ways in which, Indigenous communities have shaped forest cover, ecosystem dynamics, and species composition over the long-term, with lasting legacies for biodiversity and the carbon cycle (Pyles et al., 2022; Montoya et al., 2020; Levis et al., 2017; Pereira Cruz et al., 2020; Robinson et al., 2018). Well-known examples from the lowlands of South America include landscape-level hydrological engineering infrastructure in southwestern Amazonia (Erickson, 2000; Lombardo and Prümers, 2010; Prümers et al., 2022; Rojas Mora and Gaitán, 2015) and vast extensions of anthropogenic dark earths in the central and eastern Amazon basin (Arroyo-Kalin, 2017; Lombardo et al., 2022; Schmidt et al., 2014). Human-mediated environmental changes such as long-distance species translocations (Capriles et al., 2022; Piperno, 2011; Piperno et al., 2000), influences on plant genetic diversity (Wang et al., 2025; Lauterjung et al., 2018) modifications to vegetation composition, and alterations to fire regimes (Iriarte et al., 2020; Nascimento et al., 2022) have all been documented. In the context of the biosphere, there is also a growing appreciation of the long-term widespread and systematic promotion of useful tree species, to the extent that Indigenous agroforestry practices have left long-lasting legacy effects still detectable in modern inventories (Balée, 2013; Clement et al., 2015; Lins et al., 2015; Maezumi et al., 2018; Levis et al., 2017). However, with some notable exceptions (e.g., Levis et al., 2017), quantitative, spatial comparisons of human activities and modern biodiversity have been lacking. This is particularly the case for the Atlantic Forest, something problematic given drastic reductions in its extent and the sustainability challenges facing its ecosystems in the 21st century.

The long-term relationship between the Atlantic Forest biome and pre-Colonial Indigenous societies has been of longstanding interest (Azevedo and Scheel-Ybert, 2020; Bitencourt, 2006; Iriarte and Behling, 2007; Robinson et al., 2018; Pereira Cruz et al., 2020). By the late Holocene, multiple archaeological cultures with distinct population structures, settlement patterns, and subsistence bases coexisted in the southern Atlantic Forest (SAF). Tupi-Guarani groups were among the most numerous Indigenous peoples at the time of Conquest, represented archaeologically by the Tupi-Guarani culture, with a distinctive “Amazonian” land use pattern centered on horticulture and agroforestry along the margins of major rivers (Noelli, 2008; Bonomo et al., 2015; Iriarte et al., 2017b). In the highlands, southern Jê people practiced mixed low-level food production and agroforestry, focused especially on the management of moist highland forest dominated by *Araucaria angustifolia* (Noelli, 2005; Robinson et al., 2018). There is evidence for both demographic expansion and land use intensification over the past 2000 years, including the emergence of social organization beyond the family unit, which has been linked to an increasing anthropogenic footprint on environments (De Souza and Riris, 2021; Iriarte et al., 2017a, 2008; Reis et al., 2014, 2018). Cultural

trajectories on the Atlantic coast differ somewhat, with no clear signs of demographic expansion or land use intensification from 2,000 cal BP (Toso et al., 2021). Southern Jê contact occupations on the coast likely commence around 1,300 cal BP (Cardoso et al., 2024), while Tupi-Guarani occupations commence after 1,000 cal BP (Bonomo et al., 2015). Changes to the extent and, to a lesser extent, composition of the SAF are broadly coeval with these cultural trajectories (Behling et al., 2004; Iriarte and Behling, 2007; Gessert et al., 2011; Iriarte et al., 2017a; Jeske-Pieruschka and Behling, 2012). However, there is a general lack of systematic analysis as to how different elements of flora were selected for by communities over the long-term, how the strength of anthropic influences varied, and whether legacy effects remain detectable in modern inventories, with some notable exceptions (Robinson et al., 2018; Lauterjung et al., 2018).

It is clear that the Atlantic Forest and its resources were important to human subsistence from the earliest occupations in the region by humans (Dias, 2012). Quantitative and systematic analysis of long-term interactions between forests and people need to be developed in order to better estimate how human societies and their environments co-evolved (Nascimento et al., 2024), as well as to better comprehend modern biodiversity contexts, dynamics, and concerns. Given the complexity of Neotropical ecological communities, and indeed, socio-ecologies, initiatives that seek to foster resilience and restoration such as the Pact for the Restoration of the Atlantic Forest (<https://www.pactomataatlantica.org.br/>) can benefit from the integration of such data (Flores et al., 2024; Witteveen et al., 2025). The deficit of precise knowledge on the effects of Indigenous activity also has consequences for how initiatives and roadmaps are defined in the present; restoration targets are typically set with reference to Colonial-era or 20th century evidence on past environmental conditions (Carlucci et al., 2021). As biodiversity outcomes are enhanced by the promotion of Indigenous land use strategies (Benzeev et al., 2023), the potential of palaeodata to contribute or even suggest alternatives to modern land use patterns is clear (Silva et al., 2022). Furthermore, despite awareness that reliable baselines and consistent estimates of the strength of legacy effects on forest composition are necessary for designing impactful policies (Gillson and Marchant, 2014; Ledru et al., 2016), the exact definition of these terms tends to vary on a case-by-case basis. For our purposes we define a legacy effect as a measurable, statistically significant influence of Amerindian derived data on the performance of a plant species model.

To enhance the accuracy and dimensionality of high-level synthetic works and to generate new insights into the scope of pre-Columbian impacts and their legacies in the Atlantic Forest, the first phase of our research comprised developing two presence-only species distribution models (SDMs) using Maxent (Phillips et al., 2017; Kass et al., 2021). These models are trained on legacy archaeological site location data for the Tupi-Guarani and southern Jê Amerindian traditions (c. 2,200–500 cal. BP), compiled from the Sistema Integrado de Conhecimento e Gestão (<https://sicg.iphan.gov.br>) operated by the National Institute of Historic and Artistic Heritage of Brazil (IPHAN) and extant archaeological site compilations (Bonomo et al., 2015; Riris and de Souza, 2021). A total of 14 environmental covariates were used as predictors and Target Group Sampling (TGS) was used

to generate a distribution of background points that somewhat mitigated the spatial clustering of presences caused by sampling bias. Our approach builds significantly on similar previously published work (Pereira Cruz et al., 2020) by increasing the number of covariates used as predictors, adopting a finer modeling resolution, using more sophisticated methods for background point generation, introducing a more robust, spatially-explicit cross-validation strategy, and focusing on evaluating models using metrics appropriate for presence-only models. Next, we used the same workflow and predictors to develop 35 additional Maxent SDMs, trained on geolocated plant species occurrence data obtained from the Global Biodiversity Information Facility (GBIF), comprising mostly species of known Indigenous cultural and/or medicinal importance. Finally, we explored methods of reclassifying continuous prediction data from all models into meaningful categorical classes using Predicted/Expected (P/E) curve characteristics.

Following evaluation of performance and interpretation of covariate contribution of phase one models, the second phase of our research involved introducing reclassified Amerindian SDM predictions and Euclidean distances from Amerindian sites as additional covariates (known as Amerindian derived covariates, ADCs) within additional runs of a subset of 11 plant species models. To ensure any conclusions around the effects of introducing ADCs into plant species models were robust, we refitted these phase two models 50 times using different distributions of background points. Finally, we compare plant species models incorporating ADCs with those fitted to purely environmental covariates using Student's *t*-tests on distributions of key evaluators of model performance, such as Continuous Boyce Index (CBI) and area under the operating curve (AUC_{ROC}). Our results demonstrate that the introduction of reclassified southern Jê predictions as additional covariates significantly improves the predictive accuracy of *Araucaria angustifolia* and *Tillandsia stricta* models on unseen data, although in the latter case the predictions are less robust. These findings further indicate that prehistoric human populations left a lasting impact on the vegetation of the southern Atlantic Forest, and demonstrate that floristic ecological models can be improved by the explicit inclusion of archaeological data.

2 Materials and equipment

2.1 Data acquisition and pre-processing

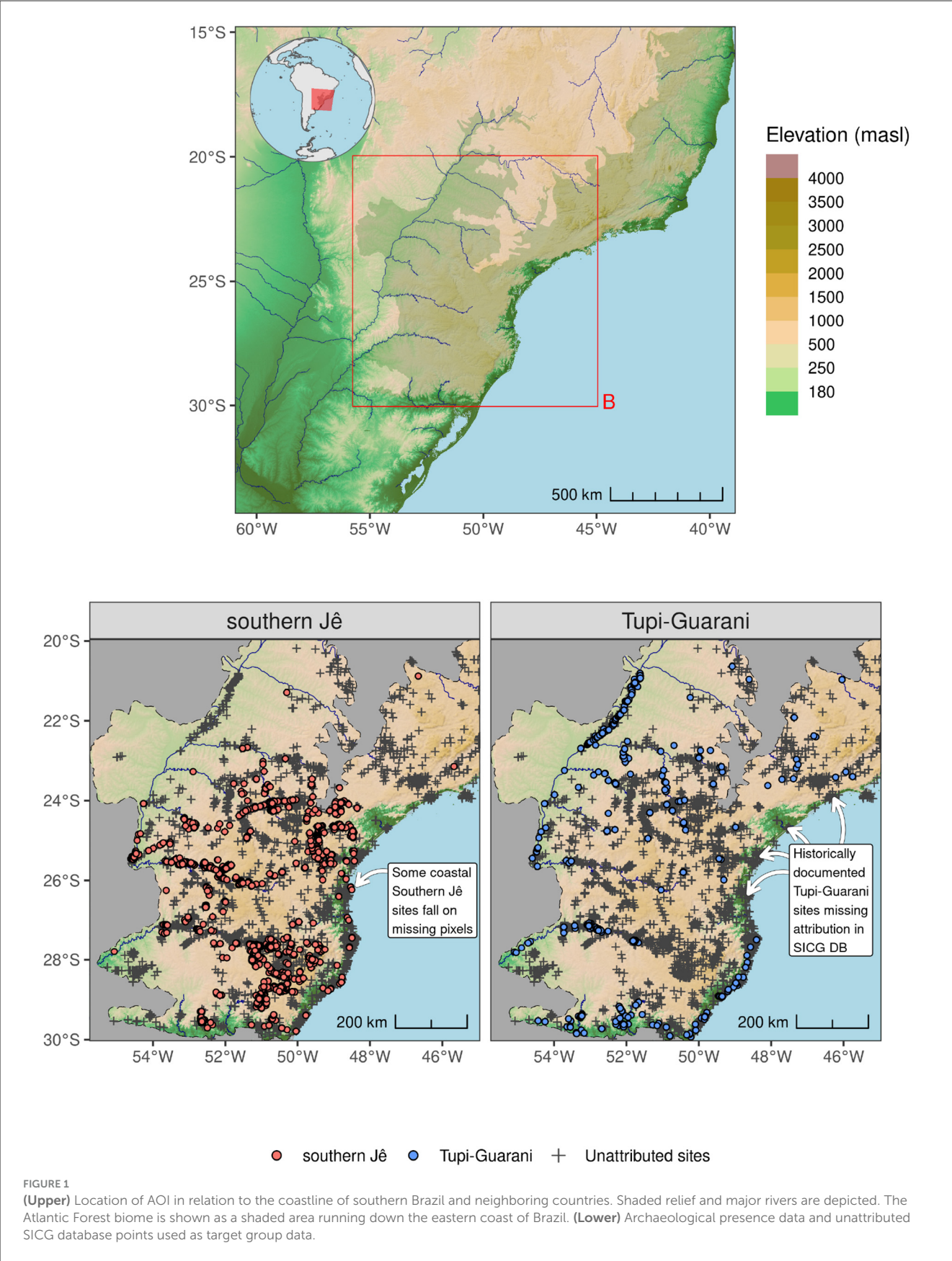
2.1.1 Archaeological site location data

This study evaluates the contemporary environmental impact of two pre-Columbian Amerindian cultures within the Atlantic Forest of southern Brazil: the Tupi-Guarani and the Southern Jê. The Area of Interest (hereafter AOI) comprises a 10 km buffer around the extent of southern Atlantic Forest (AF), where it falls between -45° and -60° longitude and -30° and -20° latitude (Figure 1) (see discussion around sectorization in Marques et al., 2021). Amerindian site locations within the AOI were collated from a variety of sources, including previously published studies (e.g., Riris and de Souza, 2021; Riris, 2019) and published databases (i.e. SICG [<https://sicg.iphan.gov.br/>], see Figure 1). Site locations

obtained from the SICG database that lacked cultural attribution data were excluded from the presences data set but retained for use in target group sampling (TGS) (see Section 3.2). These data are a valuable approximation of the locations inhabited by pre-Columbian groups, however—like all archaeological data—they are subject to systematic recovery biases resulting in specific spatial patterns that impact the modeling procedure described below. In particular, the coastal region remains difficult to reliably model both due to a lack of securely attributed archaeological sites and gaps within the environmental covariate data arising from the dynamic nature of coastal zones (see Figure 1, lower).

Archaeological evidence indicates that the Southern Jê were extant in the southern Brazilian highlands since at least 2200 calendar years before present, and were responsible for the production of Taquara/Itararé tradition ceramics. Over time, southern Jê groups developed distinctive forms of domestic and ceremonial earthen architecture, in the form of pit house dwellings and enclosed funerary mound complexes (Iriarte et al., 2013). The former can occur in clusters of more than 100 individual pits, potentially with defined trackways and terracing spatially associated to them. Pit houses are more common above ~ 800 masl; other types of settlement sites predominate at lower elevations. Funerary mounds characteristically occur in pairs of unequal size, topographical prominence, and degree of elaboration. They date from approximately $\sim 1,000$ cal BP, and are interpreted as the architectural expression of emerging political inequalities centered on the duality of historically documented southern Jê moieties (Iriarte et al., 2008; Corteletti, 2013). Larger ceremonial centers may have acted to socially integrate Jê groups on a regional scale, potentially indexing the development of supra-kinship levels of organization, with local community leaders presided over by a paramount chief. Southern Jê groups likely practiced a mixed form of agroforestry that incorporated both domesticates and managed forest resources, in particular the seeds (Portuguese: pinhão) of *Araucaria angustifolia* where this species occurs (Bitencourt, 2006; Iriarte and Behling, 2007; Robinson et al., 2018). Maize pollen is first attested in sedimentary records from areas occupied by southern Jê at $\sim 1,800$ cal BP (Gessert et al., 2011). Archaeobotanical data from excavated domestic contexts has revealed evidence of maize, squash, beans, and manioc from ~ 600 cal BP in the highlands (Corteletti et al., 2015), and maize and sweet potato from ~ 1300 cal BP on the Atlantic coast, coeval with the reoccupation of sambaqui (shell midden) sites by these peoples (Wesolowski et al., 2010; Toso et al., 2021). Evidence of maize consumption, likely in the form of chicha (beer), has also been recovered from contexts associated with funerary mounds dating from ~ 700 cal BP (Iriarte et al., 2008). A demographic transition has been suggested to coincide with the expansion of *Araucaria* forest in the highlands from about ~ 1300 cal BP as well, related to an overall intensification of land use that included fire setting and forest management, as well as increasing social complexity (Bonomo et al., 2015; Iriarte et al., 2017a; Robinson et al., 2018; De Souza and Riris, 2021).

The Tupi-Guarani (inclusively comprising the predecessors of both Guarani and Tupinambá people) appear in our study area from approximately 2,000 cal BP, somewhat later than the southern Jê (Bonomo et al., 2015; Iriarte et al., 2017b). Sites with



distinctive Tupi-Guarani tradition brushed, corrugated, and/or painted ceramics tend to concentrate along major rivers and their tributaries. Both ethnohistoric and archaeological evidence indicate intense occupations associated to anthrosol formation in Tupi-Guarani sites, representative of a period of significant population expansion originating in Amazonia during the late Holocene (Brochado, 1984; Noelli, 1998; Iriarte et al., 2017b). Large, palisaded villages were centered on plazas surrounded by extended family longhouses, which at a regional scale were organized into multi-village “confederacies” (Noelli, 1993; Milheira and DeBlasis, 2014). Villages exploited an extensive hinterland of managed forest, swidden plots, and house gardens, comprising a spatially extensive agroforestry pattern of land use that was somewhat tethered to watercourses. Like the southern Jê, Tupi-Guarani groups cultivated manioc, beans, maize, squash, yams, sweet potato, as well as peanuts (Noelli, 1993; Behling et al., 2005). The primary mode of expansion of the Tupi-Guarani archaeological culture is thought to be periodic fissioning into daughter villages, with each successive settlement incorporating a similar hinterland. By the European conquest, Tupi-Guarani groups were present throughout our entire study area, with extensive settlement along the coast. As such, they were among the first Indigenous people encountered and impacted by Portuguese colonization.

Due to variation in the geographical distribution, organization, land use patterns of these archaeological cultures, and the affordances of highland vs. riparian environments, we anticipate concordant differences in associations between pre-Columbian occupations and plant species, as represented in our presence datasets. Nevertheless, iconic keystone species like *Araucaria angustifolia* notwithstanding (Robinson et al., 2018; Lauterjung et al., 2018), there is a lack of systematic evidence on which floristic elements may have been preferentially interacted with, how the strength of anthropic influences varied, and whether legacy effects remain detectable in modern inventories. Beyond the selection of plant presence records detailed below, we do not presume the presence or absence of any association.

2.1.2 Floristic occurrence data

All Plantae presence records with a country classification of “Brazil” with coordinates (~11.2M) were obtained from the Global Biodiversity Information Facility (GBIF) database (GBIF.org, 2024). These data were spatially filtered to leave only those presences falling within the extent of the Atlantic Forest biome buffered by c. 5 km within the AOI (see Figure 1). Records were further filtered by attribute to include only those with presences identified down to species level. Records with invalid coordinates or missing observation dates were also removed. Remaining records were checked against the Artificial Hotspot Occurrence Inventory (AHOI) (Park et al., 2023) and any records matching AHOI values were removed. Finally, exact duplicates were removed as presences were reduced to a single record per species, per coordinate pair. Approximately 800k individual presence records, spanning over 25,000 unique species, remained following pre-processing.

Our initial selection for modeling comprised three sets of species: (i) the top 20 most abundant species within the processed

GBIF data; (ii) any species recorded as involved in recent Mbyá-Guarani medicinal plant exchanges in southern Brazil (de Andrade et al., 2021) and other key culturally-important species, such as *Araucaria angustifolia* (see Behling et al., 2004; Iriarte and Behling, 2007; Lauterjung et al., 2018; Robinson et al., 2018) also represented in the GBIF data, and finally iii) tree species noted as being particularly important in Indigenous agroforestry systems: guabiroba (*Campomanesia xanthocarpa*), guabiju (*Myrcianthes pungens*), araçá (*Psidium* sp.), goiaba (*Acca sellowiana*), and comestible palms (*Eurpepe* sp. and *Syagrus romanzoffiana*) (Cassino et al., 2021; Corteletti et al., 2015). 18 species with less than 100 occurrences were removed, though one species (*Myrcianthes pungens*) with 77 presences was retained due to its importance within agroforestry. Records relating to *Hedychium coronarium*, *Ipomoea cairica*, and *Rhaphiolepis bibas* (the former two introduced to the Americas from Asia, the latter from Africa) were flagged as introduced species and not progressed to modeling. Initial exploratory models also highlighted 9 poorly performing species, presence data relating to which were also removed. After pre-processing, ~33,000 presence records spanning 35 unique plant species remained. Finally, a number of botanical databases (POWO, 2024) and relevant published ethnobotanical literature (Martínez Crovetto, 2012; Keller et al., 2010; Bueno et al., 2005) were consulted to confirm and/or ascertain traditional medicinal or cultural uses of all modeled plant species.

An important element that warrants future consideration, but which we do not explicitly address here, is the time-dependence of plant occurrences within the GBIF database. The legacies of more recent land use practices (such as mining, logging, ranching, especially following industrialization and mechanization) are likely to have a more dramatic effect on species distribution and forest composition than Indigenous land use, which in many cases is removed by several centuries or more. Although biased toward the present day, especially after 1950, the GBIF database also contains records from before this, extending in some cases to the nineteenth century. We have chosen to treat these data as a first-order approximation of the distribution of the modeled species, and anticipate that in future work historical observations such as naturalist accounts of the Atlantic Forest, and other Colonial-era writings such as land surveys, will be instrumental in closing this known gap in data. Palaeoecological records may also serve as a useful point of comparison and potential check on model results. It must, however, be acknowledged that the overall effect of land use from the Colonial period onwards likely had a degrading effect on species distribution, and our use of this data likely underestimates the historical distribution.

3 Methods

3.1 Feature selection and data pre-processing

Modeling resolutions of approximately 1 km, 5 km, and 10 km were adopted and, at each resolution, presence data were spatially filtered so only a single presence location per modeled pixel remained (also known as spatial-grid thinning or rarefaction). Cumulative distance rasters were produced from each set of

occurrence data for use as model-specific covariates (see [Allouche et al., 2008](#)). Following spatial filtering at 1 km, the number of archaeological presence data comprised 349 Tupi-Guarani and 1,136 Southern Jê records. Plant species with less than 100 occurrences at 1 km were not progressed to modeling, and the final number of plant species presences ranged from 77 (*Myrcianthes pungens*) to 1,186 (*Psidium guajava*). Initially, 172 environmental covariates were obtained as raster data of varying resolutions from ten open repositories (see [Supplementary Table S2](#)). These covariates described the recent edaphic, bioclimatic, phenological and hydrotopographic characteristics of the landscape ([McDowell et al., 2023](#); [Johnson et al., 2019](#); [Fischer et al., 2008](#); [Lehner and Grill, 2013](#); [South et al., 2023](#); [Joint Research Centre, 2022](#); [Wieder et al., 2014](#); [Poggio et al., 2021](#); [Fick and Hijmans, 2017](#); [Amatulli et al., 2018](#)).

The use of recent covariate data to fit models of pre-Columbian archaeological site distributions, which were established under earlier bioclimatic conditions, risked reducing model performance but was a necessary compromise that allowed for both the wide temporal span of the archaeological sites and the relatively recent provenance of plant occurrence observations (though sub-optimal from a modeling perspective, correlation between historic and modern covariates means this approach has proved effective for e.g. locating new archaeological sites in the Amazon, see [Walker et al., 2023](#)). The use of palaeoclimatic covariates for archaeological models was also precluded by the necessity to maintain consistency of covariates between modern plant and archaeological models, so that the explanatory power of introducing Pre-Columbian presence probability data into modern plant species distribution models could be fairly assessed (see Section 4.4).

From the initial collection of 172 covariates, a subset of 14 generically-performant candidate covariates (see [Figure 2](#)) were identified by first selecting 60–70 covariates showing suitable variance across presence and background points, per species. Collinearity between regression model predictors can cause increases in the variance of the regression coefficients, making them unstable and difficult to interpret ([Akinwande et al., 2015](#)). Accordingly, pairs of covariates were assessed for high collinearity (see [Figure 3](#)) using Pearson's correlation coefficients and pairs with values over 0.6 were addressed by removing whichever covariate ranked lower on a preference list comprising covariates already established as important to modeling key plant species and Amerindian groups within the Atlantic Forest, or subtropics more widely (see [Iriarte and Behling, 2007](#); [Rafuse, 2021](#); [Walker et al., 2023](#)), such as median elevation, distance to coastline ([Pereira Cruz et al., 2020](#)), aspect ([Robinson et al., 2018](#)) and height above nearest drainage ([Levis et al., 2017](#)). Further feature processing was undertaken (see [Supplementary Data Sheet 1](#) for additional details).

3.2 Generating background points

In addition to presence points, Maxent models also require 'background points', which—distinct from absence or pseudo-absence points—are generated locations positioned to capture the environmental conditions available to the species being modeled.

The model then uses presences and background points to distinguish between suitable and less suitable habitats, respectively ([Sillero and Barbosa, 2021](#): 214–5). It is important to mitigate the spatial bias within presence records when designing SDMs (see [Valavi et al., 2022](#); [Boria et al., 2014](#); [Phillips et al., 2009](#); [Steen et al., 2021](#); [Kramer-Schadt et al., 2013](#)). Target Group Sampling (TGS) was used as this has been shown to be effective for ecological modeling ([Baker et al., 2022](#); [Barber et al., 2022](#)). For archaeological presence data, the target points comprised 7,412 unattributed sites locations obtained from the SICG database. For GBIF species presence data, the target points comprised presences for all other species to those being modeled, which ranged from 21,354 to 22,682 points. For each archaeological tradition or plant species, the number of background points generated was 10,000, an arbitrary threshold which has been shown to perform well in a number of contexts ([Whitford et al., 2024](#); [Valavi et al., 2022](#)). Preliminary models performed better when background points were not sampled from presence pixels or pixels immediately adjacent to presence pixels, so these pixels were removed from KDEs prior to background point sampling (see [Figure 4](#)).

3.3 Model specification (Maxent) and cross-validation

Maxent is a good choice for an initial exploration of the relationship between Amerindian land-use and modern floristic composition as it is an accessible ([Phillips et al., 2017](#)), well-tested, generically-applicable (i.e. suitable for both human and plant species data) modeling technique ([Júnior and Nóbrega, 2018](#); [Whitford et al., 2024](#)) that consistently benchmarks above many other techniques when used for presence-only (presence-background) data ([Valavi et al., 2022](#)). It is also less sensitive to the number and prevalence of presences relative to other techniques e.g., random forests ([Grimmett et al., 2020](#)) and has been more widely used within archaeology and palaeoecology relative to other modeling approaches ([McMichael et al., 2014](#); [Pereira Cruz et al., 2020](#); [Rafuse, 2021](#); [Howey et al., 2016](#); [Demján et al., 2022](#); [Wachtel et al., 2018](#)). To maximize the independence of test and training partitions and reduce the chance of overly optimistic evaluation scores ([Roberts D. R. et al., 2017](#)), a 7-fold spatial cross-validation routine was adopted using a partition block size determined using the R package {blockCV} ([Valavi et al., 2019](#)) (see [Supplementary Data Sheet 1](#) for further information). Relatively simple parameter sweeps using feature classes and regularization values were specified and aimed to provide enough flexibility to effectively capture and model individual species' distributions whilst reducing the chance of overfitting, which can be especially problematic when comparing predictions from many models ([Morales et al., 2017](#); [Merow et al., 2013](#)).

4 Results

The Amerindian model predictions produced by this study represent, to our knowledge, the most comprehensive predicted relative likelihood of occurrence maps for Southern Jê and Tupi-Guarani settlement (c. 2,200–500 cal. BP) in the southern

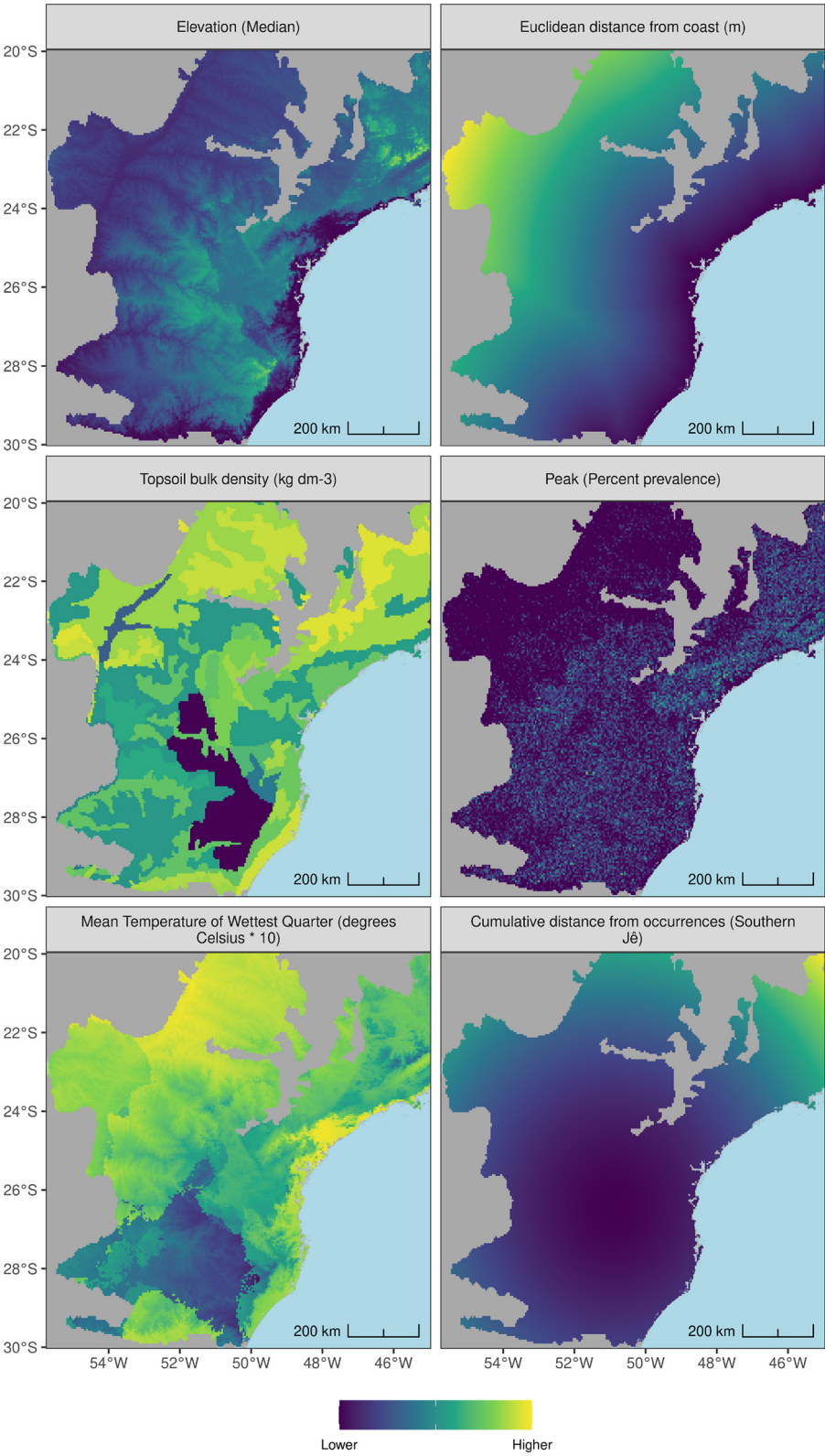
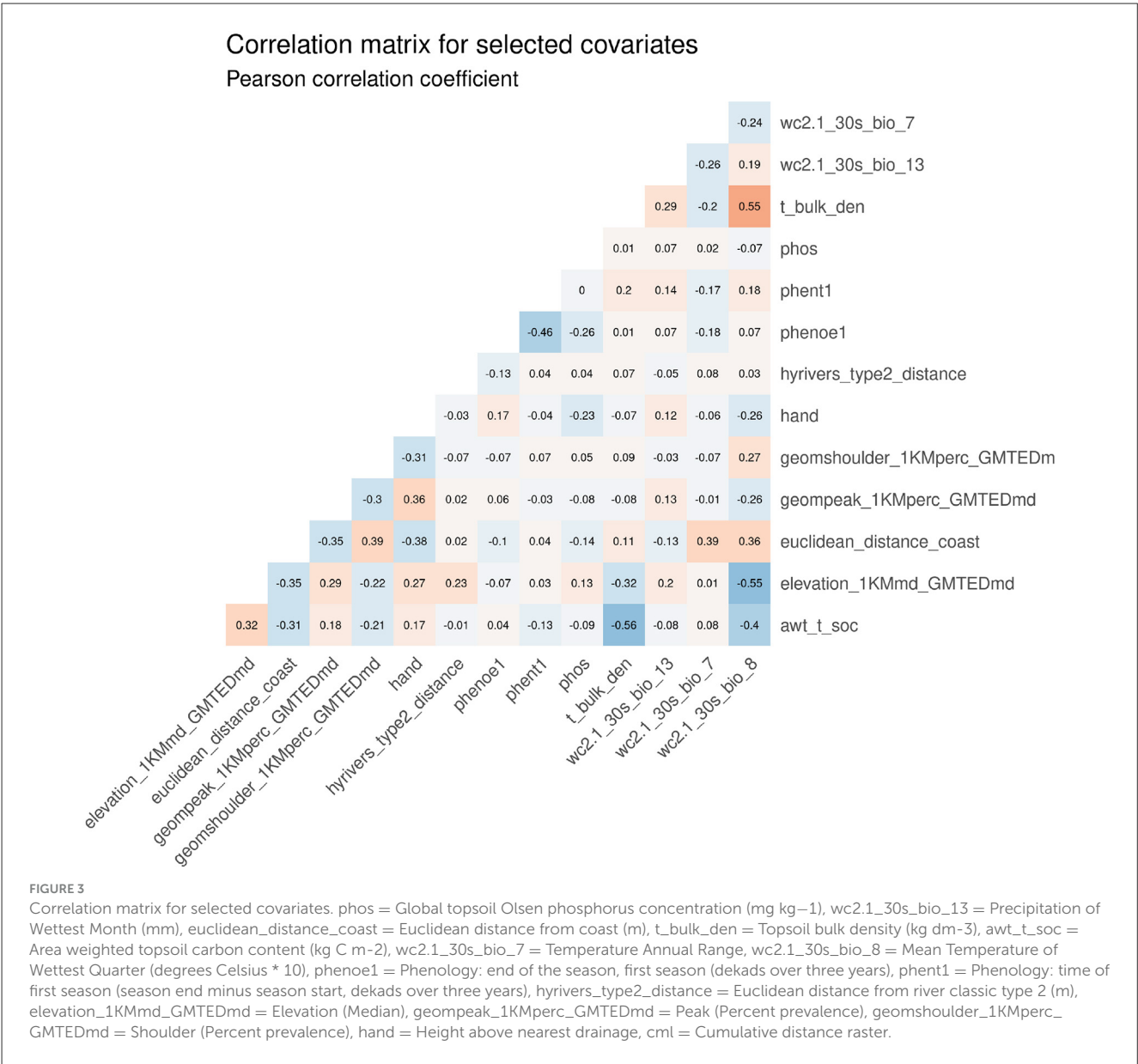


FIGURE 2
Pre-processed covariates obtained for the AOI from various sources. Elevation (Median): GMTED; Euclidean distance from coast (m): Natural Earth Hi Res; Topsoil bulk density (kg dm-3): Regrided Harmonized World Soil Database v 1.2; Peak (Percent prevalence): GMTED; Mean Temperature of Wettest Quarter (degrees Celsius * 10): WorldClim version 2; Cumulative distance from occurrences (Southern Jê): User generated.

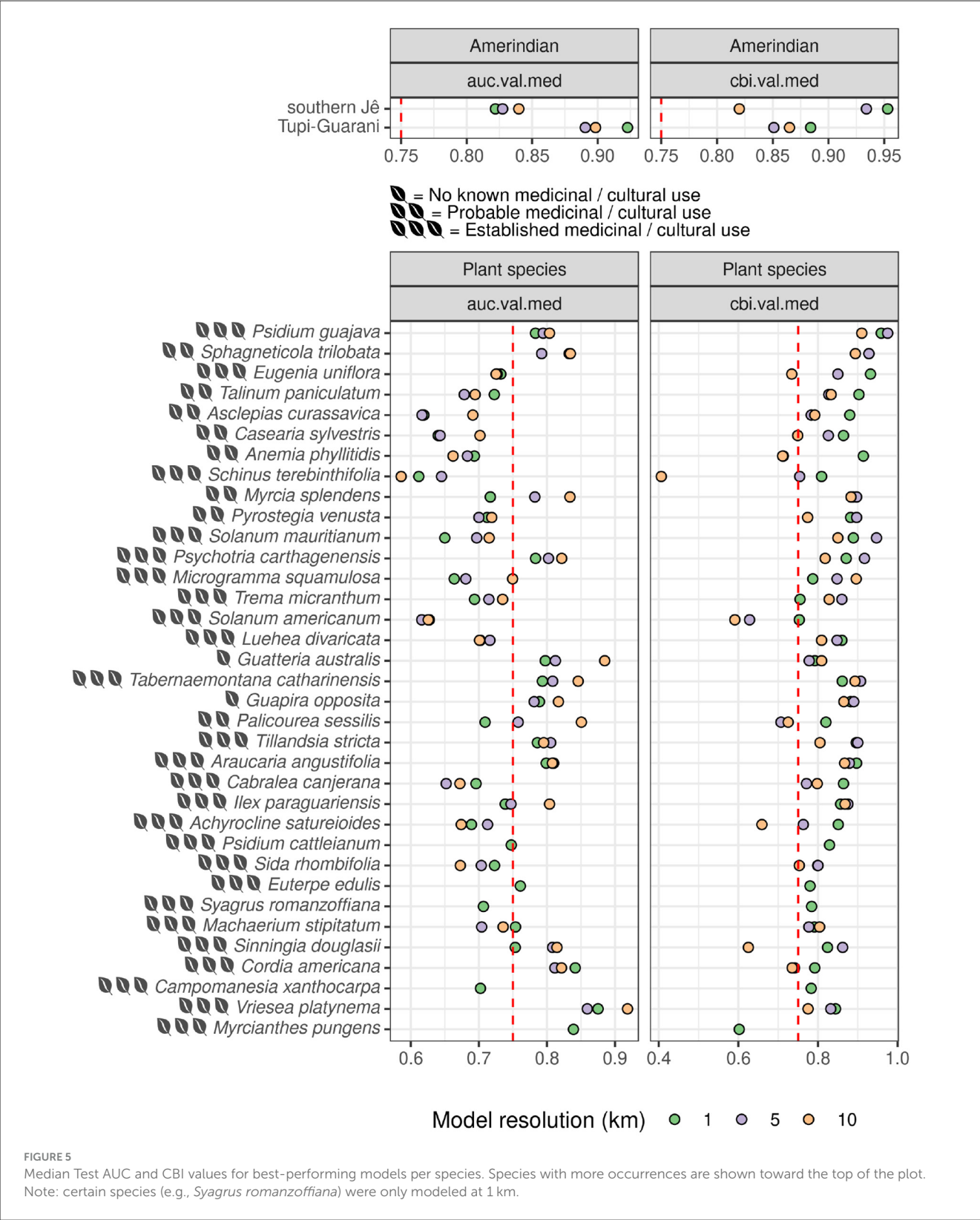


Atlantic Forest to date. These models build on and improve previously published models (Pereira Cruz et al., 2020) both in terms of spatial resolution and performance across multiple evaluators (see Figure 5 below). In terms of model performance, AUC (discussed further below) assesses the ability of models to discriminate presences from absences by expressing the ratio of wrongly predicted absences (false negatives) to correctly predicted presences (true positives). It ranges from 0–1, with 1 representing perfect discrimination and 0.5 indicating a rate no better than random. The 10 km-resolution Southern Jê and Tupi-Guarani models produced by this study achieved average AUC_{Test} values of 0.81 and 0.88, when tuned to maximize this evaluator (see Supplementary Table S4). These values represent improvements of 0.10 and 0.03, respectively, when compared to previously published Maxent models of the same Amerindian traditions (i.e., Pereira Cruz et al., 2020). However, using AUC for presence-background data is problematic (see Barbet-Massin et al., 2012; Liu et al., 2013;

Jiménez and Soberón, 2020) and a more appropriate evaluator—the Continuous Boyce Index (CBI)—is therefore reported (see Boyce et al., 2002; Li and Guo, 2013, p. 795; Jiménez-Valverde, 2012; Sillero and Barbosa, 2021; Lobo et al., 2008; Yackulic et al., 2013; Leroy, 2023). CBI measures the correlation between predicted and expected frequencies of presence points based on area, ranging from ±1, with 0 describing a model that correctly predicts presences at a rate no better than chance (see Hirzel et al., 2006; Whitford et al., 2024). 1 km-resolution Southern Jê and Tupi-Guarani models achieved impressive average CBI_{Test} values of 0.9 and 0.81, respectively and exhibited P/E curves indicating their predictions were robust and of high-resolution.

Plant species model performance was more mixed, depending on the extent to which the chosen environmental covariates captured occurrence distribution and the frequency and distribution of occurrence data available (i.e., CBI_{Test} increased broadly in line with the number of presences available). At 1





km-resolution, the *Araucaria angustifolia* model performed strongly (average AUC_{Test} and CBI_{Test} of 0.78 and 0.85, respectively) and its spatial predictions broadly match and enhance those previously published (i.e., Marchioro et al., 2020). Particularly performant plant species models also included *Sphagneticola trilobata* (average AUC_{Test} and CBI_{Test} of 0.8 and 0.91, respectively) and *Psidium guajava* (average AUC_{Test} and CBI_{Test} of 0.77 and 0.93, respectively). Disregarding AUC as

a reliable indicator (see Whitford et al., 2024), several other models performed exceptionally well according to average CBI_{Test} alone, including *Eugenia uniflora* (0.86) and *Anemia phyllitidis* (0.86) and *Psychotria carthagenensis* (0.85). Covariate importance plots (Figure 6) and partial response plots (Figure 7) elucidate the similarities and differences between habitats favored by Pre-Columbian Amerindian traditions and a number of key, culturally- and medicinally-important plant species, helping to better characterize and quantify the extent of environmental niche overlap often cited as a limiting factor in understanding the impact of archaeological settlement on modern vegetation patterns in South America (Robinson et al., 2018; Pereira Cruz et al., 2020; Levis et al., 2017).

Several trends are discernible from the evaluation data shown in Figure 5: (i) finer model resolution (i.e., 1 km) generally improved model performance (see Manzoor et al., 2018) but only when measured according to median CBI_{Test}. The relationship between resolution and performance was more heterogeneous when measured according to AUC_{Test}, with models produced at 10 km frequently ranking as the best performing per species (though this was not the case for either Amerindian model); (ii) the 1 km Amerindian models were some of the best performing models overall, with Tupi-Guarani and Southern Jê models achieving median AUC_{Test} scores of ~0.9 and ~0.85 and median CBI_{Test} scores of ~0.89 and > 0.9, respectively; (iii) A number of plant species models fall below the commonly-cited acceptable AUC_{Test} threshold of > 0.75, suggested by Elith et al. (2006): 137 as a “useful” amount of discrimination [though 0.7 is another threshold suggested as meaningful by Pereira Cruz et al. (2020): 13]. Several of these models are from species with high numbers of initial occurrences (e.g. *Eugenia uniflora*, *Talinum paniculatum*, *Casearia sylvestris*). Overall, far fewer models fall below 0.75 when measured using CBI_{Test}, a threshold suggested by Hirzel et al. (2006: Figure 5) as representative of a “good” model, though there has been far less discussion around what constitutes an acceptable threshold for CBI_{Test} in comparison to AUC_{Test}.

4.1 Covariate importance

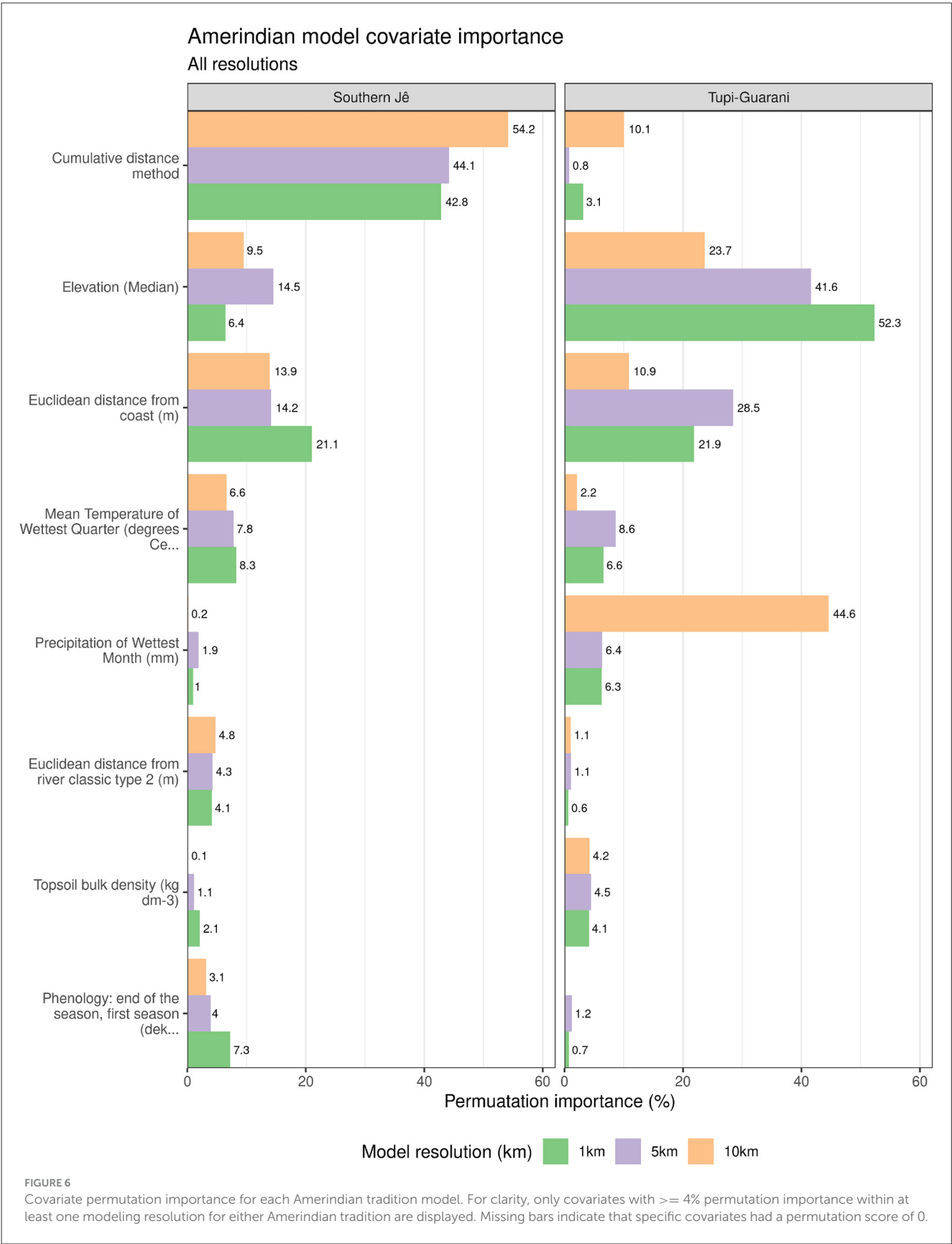
The model with the highest median CBI_{Test} value—as the most reliable evaluator of predictive performance—per species was selected to examine covariate importance at each modeling resolution. Covariate importance was assessed using both percentage contribution and permutation importance, though a greater focus is placed on permutation importance as a more stable measurement. Median elevation ranked as the most important covariate for the Tupi-Guarani model at resolutions of 1 km and 5 km, with permutation importance scores of ~52% and ~41%, respectively (Figure 6). At 10 km resolution, elevation still remained important (~23%) though was second to the climatic covariate of Precipitation of the wettest month (~44%). The high importance of elevation likely reflected the near universal position of Tupi-Guarani sites within low-lying river valleys, a well-known, archaeologically-resolved pattern (Bonomo et al., 2015; Noelli, 2004) (see Figure 7). For the Tupi-Guarani model, Precipitation of the wettest month remained an important covariate across all

resolutions, though steadily decreased in importance from coarser to finer modeling resolutions. Euclidean distance from coast was also an important covariate for the Tupi-Guarani model, ranking as the second or third most importance covariate depending on resolution, scoring ~22%, ~28% and ~10% for modeling resolutions of 1 km, 5 km, and 10 km respectively. This accords well with Pereira Cruz et al. (2020): 6, who also identified this covariate as relatively important to the Tupi-Guarani model. In contrast to the Southern Jê model (see below), the cumulative distance to all sites covariate score for the Tupi-Guarani model was comparatively low (~3%, < 1%, 10% at 1 km, 5 km, 10 km resolutions, respectively). This could be due to the uneven spatial distribution of Tupi-Guarani sites in relation to latitude, with one cluster north of 24°S and another south of 27°S, with little between. The impact of this spatial segregation may have lessened as the modeling resolution coarsened.

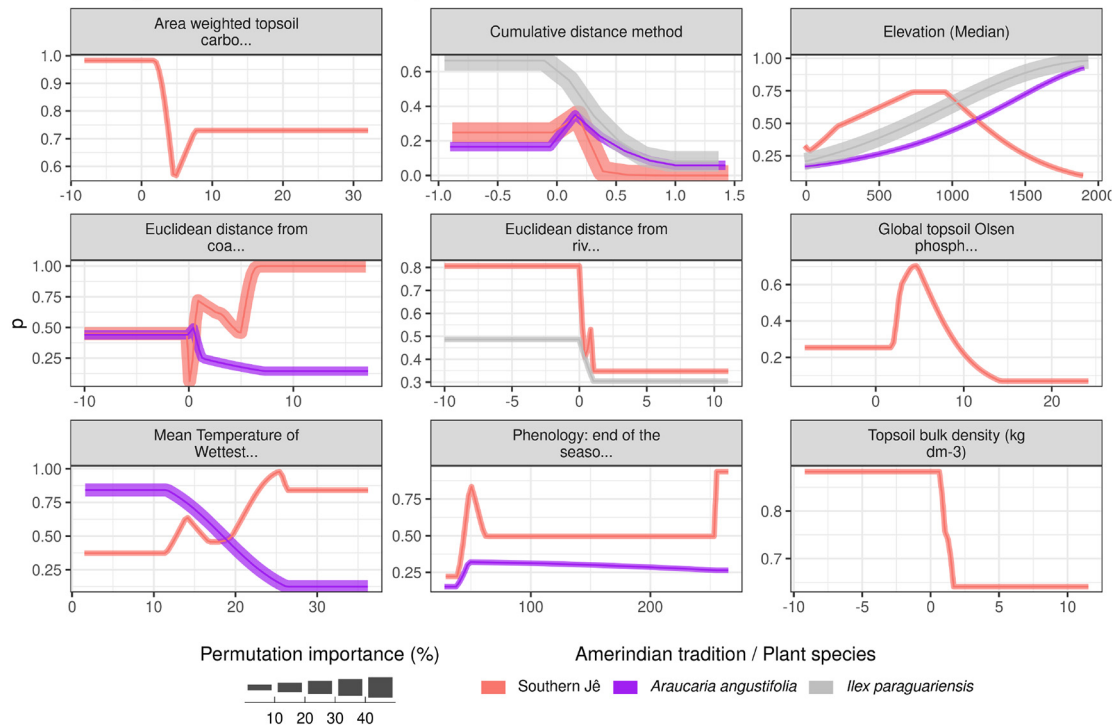
For the Southern Jê model, the most important covariate at all modeling resolutions (c. 43%–54%) was the cumulative distance to all other Southern Jê sites, indicating that Southern Jê presences were clustered and spatially-constrained in their distribution. In contrast to the Tupi-Guarani model, median elevation ranked as the second or fifth most important covariate depending on modeling resolution (~6%, ~14%, ~9% at 1 km, 5 km, 10 km resolutions, respectively). Euclidean distance from coast was consistently the second or third most important covariate (~21%, ~14%, ~14% at 1 km, 5 km, 10 km resolutions, respectively) and followed a comparable pattern some other topographic covariates in terms of losing importance as modeling resolution coarsened. These findings differ somewhat from the results of Pereira Cruz et al. (2020), who found that elevation was the highest contributing covariate to the Southern Jê model but was relatively low-ranking within the Tupi-Guarani model. This discrepancy may be because the models in this study were trained using a higher number of covariates, some of which (e.g. cumulative distance, Mean temperature of Wettest quarter, distance from coast etc.) could have acted as proxies for median elevation. The relatively low importance of Euclidean distance to rivers for either Tupi-Guarani or Southern Jê models reported in this study remains one of most marked difference with previous studies (e.g. Pereira Cruz et al., 2020). Though this may be due to differences in the river related covariates themselves [this study grouped rivers using the “classical ordering system” (Lehner and Grill, 2013) rather than Otto Pfafstetter’s hierarchy], it is again likely due to the model obtaining similar information from a different covariate(s), such as height above nearest drainage (abbrv. hand).

4.2 Response plots

Response plots illustrate model prediction values produced over a range of values for each covariate of interest, whilst holding all other covariates at median. The influence of individual covariates on model scores for each Amerindian tradition model and two culturally important species and/or those with similar environmental niche overlap is shown in Figures 6, 7 (1 km model resolution only). For the Southern Jê, Figure 7 shows that elevation values of between 500-1000 yield the highest prediction values,



Southern Jê covariate response plots
Showing the Amerindian model and two plant models.



Tupi-Guarani covariate response plots
Showing the Amerindian model and two plant models.

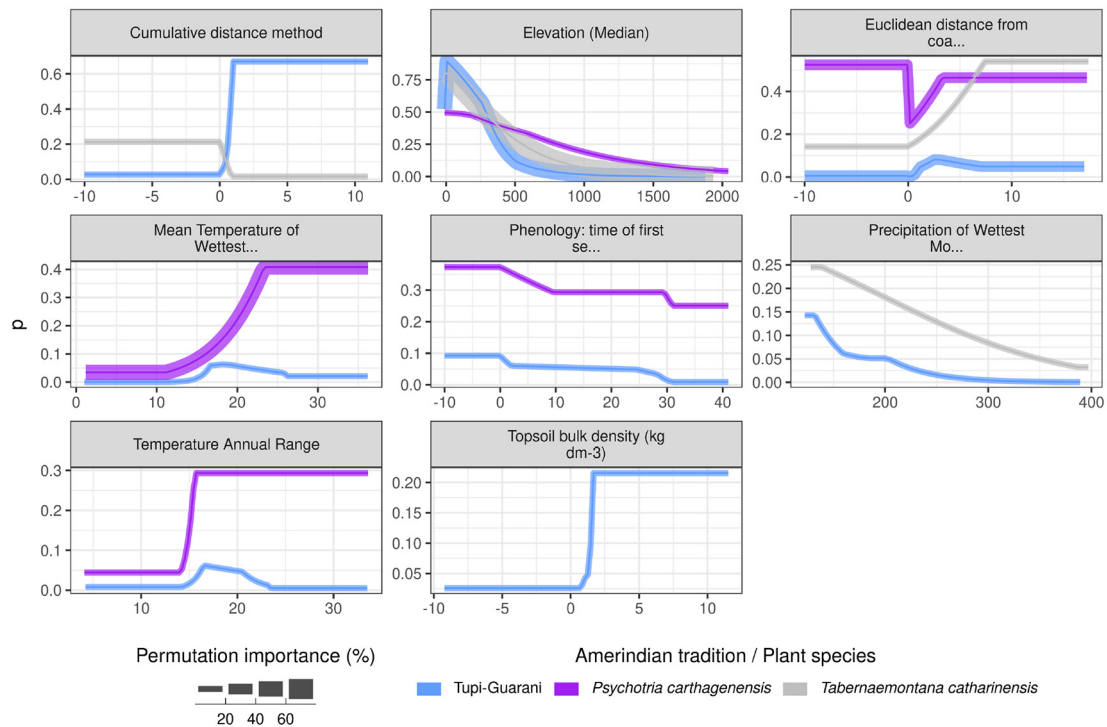


FIGURE 7
Covariate response plots. Only covariates with > 1% permutation importance to each Amerindian model are shown. Where responses are missing for plant species, this either indicates that this covariate was not included in the plant species model (e.g., eliminated due to no variation across training data) or that it had < 1% importance.

which agrees well with previously published results (Pereira Cruz et al., 2020: Figure 3), though this study indicates that elevation contributes less than 10% to the overall model at 1 km resolution. Elevation is more important for both comparator plant species shown (*Ilex paraguariensis* and *Araucaria angustifolia*) which both also favor higher elevations, with predictions increasing in line with elevation values from 500 up to 1,500 and beyond. The Cumulative distance method plot shows that the Southern Jê predictions increased with very low values (i.e. predictions were strongest close to existing sites), a pattern shared to a lesser extent by both comparator species. Mean Temperature of Wettest Month ranked as important to the *Araucaria angustifolia* model, with predictions highest at low values and steadily decreasing to zero by ~25. This accords with existing studies showing that *Araucaria angustifolia* distribution shows a preference for colder winters (Souza, 2021). The Southern Jê model diverged in this respect, with predictions increasing with higher values, though this covariate was less important to the model overall (~7%).

The response plots for the Tupi-Guarani model (Figure 5) feature responses from two plant species models that share similar environmental niches with Tupi-Guarani settlement sites (see Supplementary Figure S1): *Psychotria carthagenensis* and *Tabernaemontana catharinensis*. The greatest similarity across all models is the response to elevation (around 20–30% importance for all models), which is highest at the lowest values and gradually decreases up to 1,000, though a low-level of prediction strength for *Psychotria carthagenensis* persists at higher elevations, in contrast to the other models. Obvious differences include Euclidean distance from coast, which ranks as an important covariate (~20%) for all three models—Tupi-Guarani prediction strength increases from 0 up to around 4 before dropping off and bottoming out at around 6, whilst both plant species show a broadly linear increase in prediction strength from 0 up to 4 (*Psychotria carthagenensis*) and 7 (*Tabernaemontana catharinensis*), the maximum value within the training data.

4.3 Predictions and reclassification

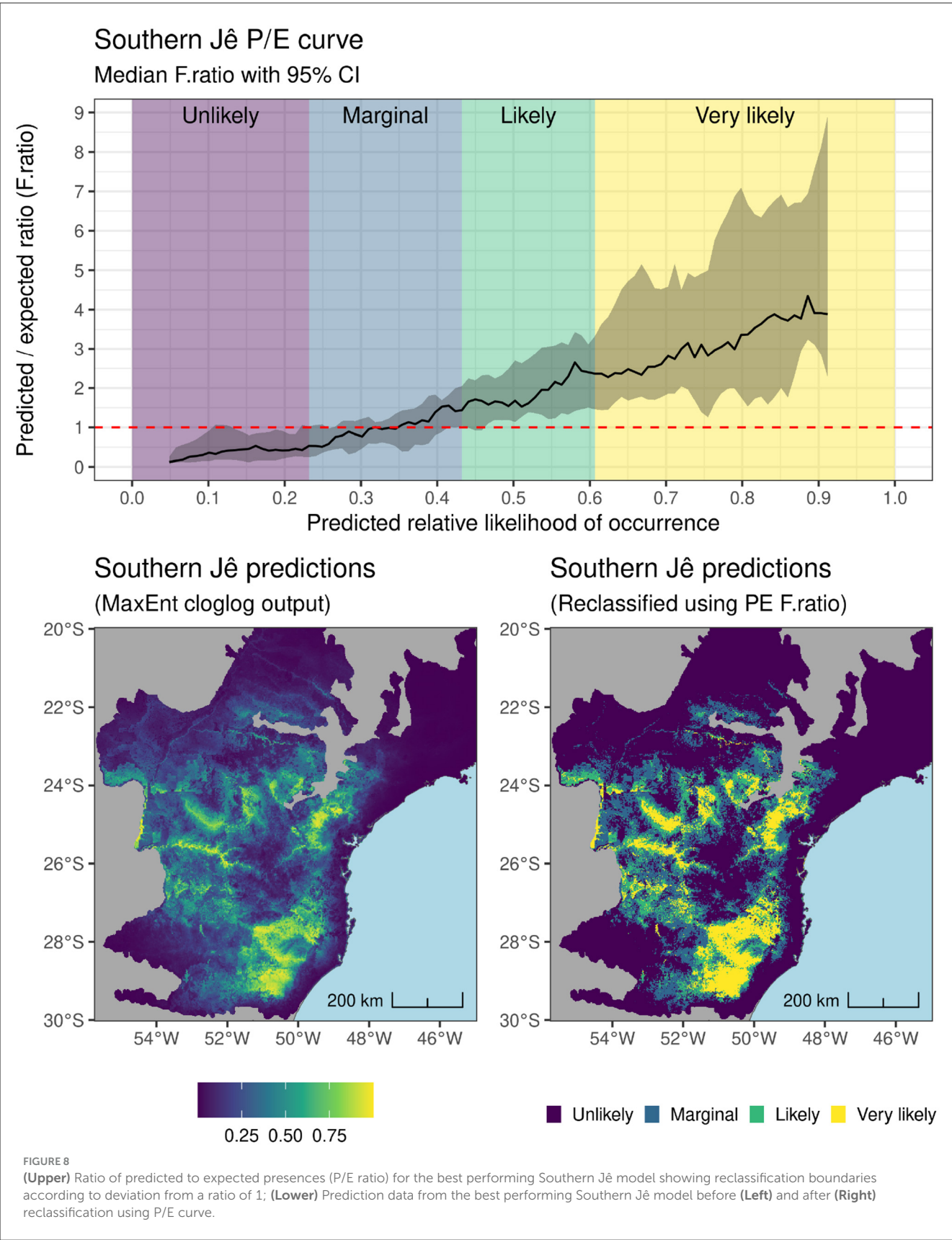
There remains some debate over exactly what SDM prediction values produced by presence-only models conceptually represent (see e.g., Fithian and Hastie, 2013), though Maxent model predictions (ranging from 0–1) continue to be reasonably interpreted as the “predicted relative likelihood of occurrence” of the modeled phenomena (Valavi et al., 2021: 1737). An example of the mapped continuous predictions (defaulting to cloglog scale in the version of Maxent used here) for the Southern Jê model can be seen on the lower left-hand side map in Figure 8. These continuous data can be analyzed directly to e.g., establish whether different modeling algorithms/covariate sets produce significantly different predictions using pairwise correlation coefficients (e.g. Bucklin et al., 2015: 22), as long as the underlying models share the same training/test data (i.e. model the same species) and geographic areas (Lobo et al., 2008; Merow et al., 2013: 10). Alternatively, continuous prediction data can be usefully be reclassified into categorical data (usually binary presence-absence maps) to facilitate interpretation. Here, the predicted to expected occurrence ratio (P/E ratio) at varying value ranges of predicted relative likelihood

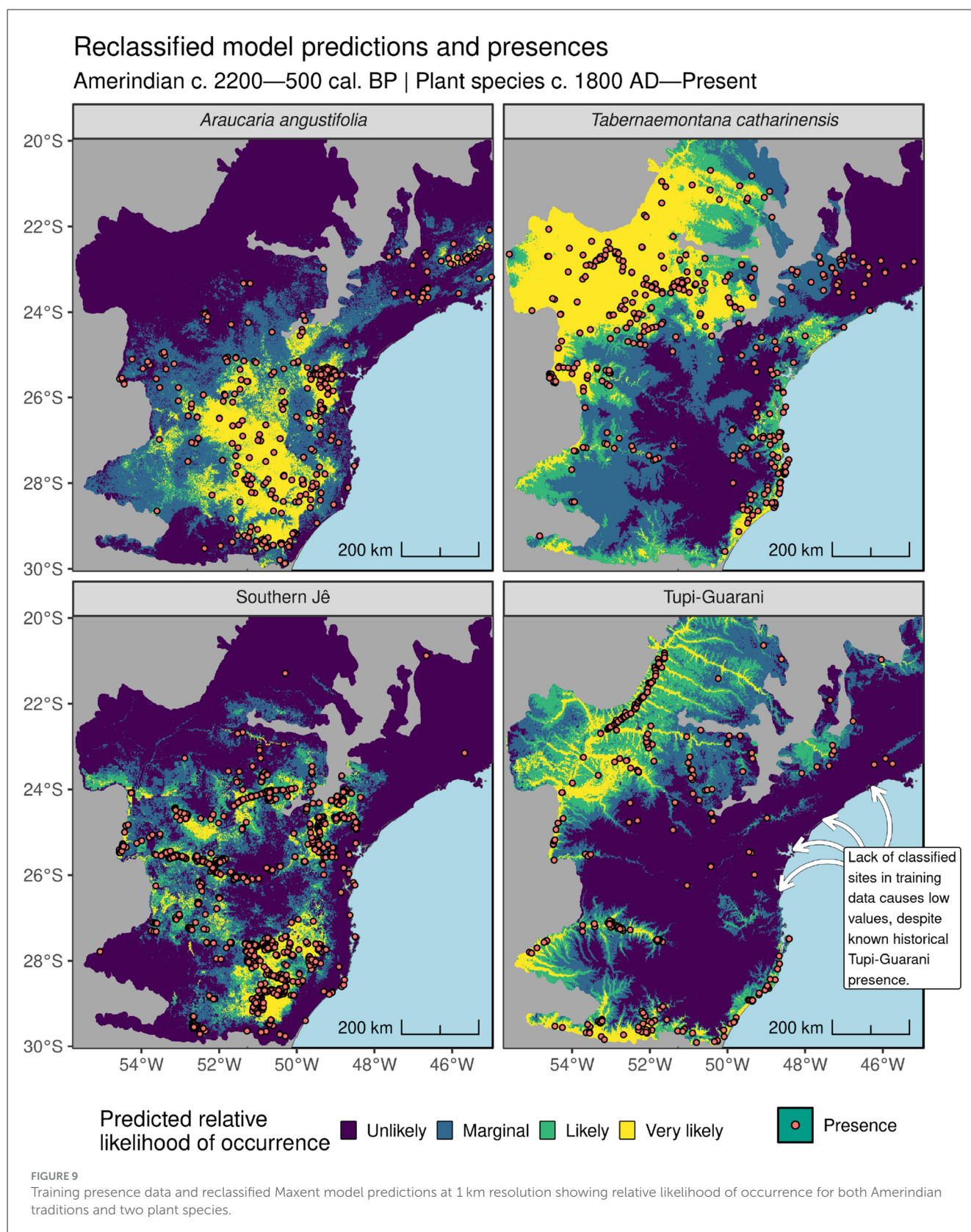
of occurrence was calculated across all spatial folds used for model cross-validation to create a P/E curve, which was used to reclassify the predictions into “Unlikely,” “Marginal,” “Likely” and “Very likely” (Hirzel et al., 2006; Yu et al., 2024; also see Figure 9). Inspection of P/E curves facilitated greater insight into model performance, such as model robustness at varying values of relative likelihood of occurrence, which can be defined by the width of the 90% CI of P/E ratios of all cross-validation folds. A narrower CI indicates a more spatially-robust model that yields similar P/E ratios across all spatial folds. As with the Southern Jê model shown in Figure 8, upper, model robustness can vary along the predicted likelihood of occurrence curve and help identify models that can more reliably predict absences than presences (indicated by the narrower CI on the left side of Figure 8, upper).

4.4 Amerindian-derived covariates (ADCs)

For the final phase of the study two sets of Amerindian-derived covariates (ADCs) were introduced into plant species models as predictors: (i) reclassified Amerindian SDM predictions for Southern Jê and Tupi-Guarani and; (ii) rasters describing the Euclidean distance from both Southern Jê and Tupi-Guarani sites (see Figure 10). To investigate the impact of ADCs, model runs both including and excluding them were bootstrapped 50 times on a subset of 11 plant species. These species included known cultural keystone species, such as *Araucaria angustifolia*, as well as tree species considered important in Indigenous agroforestry systems such as *Euterpe edulis* (Corteletti et al., 2015), and epiphytic species such as *Tillandsia stricta*. The number of available training presences relating to selected species ranged from 130–932. For each of the bootstrap runs, a different set of 10,000 background points sampled using weights from the plant species’ target group KDE (see Section 3.0.2) were generated and used to train a model using; (i) just the environmental covariates described above; (ii) the same environmental covariates plus Euclidean distances from Tupi-Guarani and Southern Jê sites and; (iii) the same environmental covariates plus reclassified Amerindian SDM predictions for Southern Jê and Tupi-Guarani. All of the ADCs were screened for high VIF and collinearity with existing environmental covariates and plant species cumulative distance rasters. All ADCs exhibited VIF scores of below 5 and both reclassified SDM predictions exhibited collinearity scores of less than ± 0.6 , with the exception of Tupi-Guarani and median elevation (-0.66). Euclidean distances from Amerindian sites were more frequently highly correlated with environmental covariates, with the Southern Jê data exhibiting positive correlations of between 0.6–0.74 with various plant species cumulative distance rasters. There is some debate around the impact of covariate collinearity on model performance (see Feng et al., 2019; Elith et al., 2011), however collinearity can impact reported covariate importance scores, so caution must be exercised when interpreting these data from models using Euclidean distances from Amerindian sites.

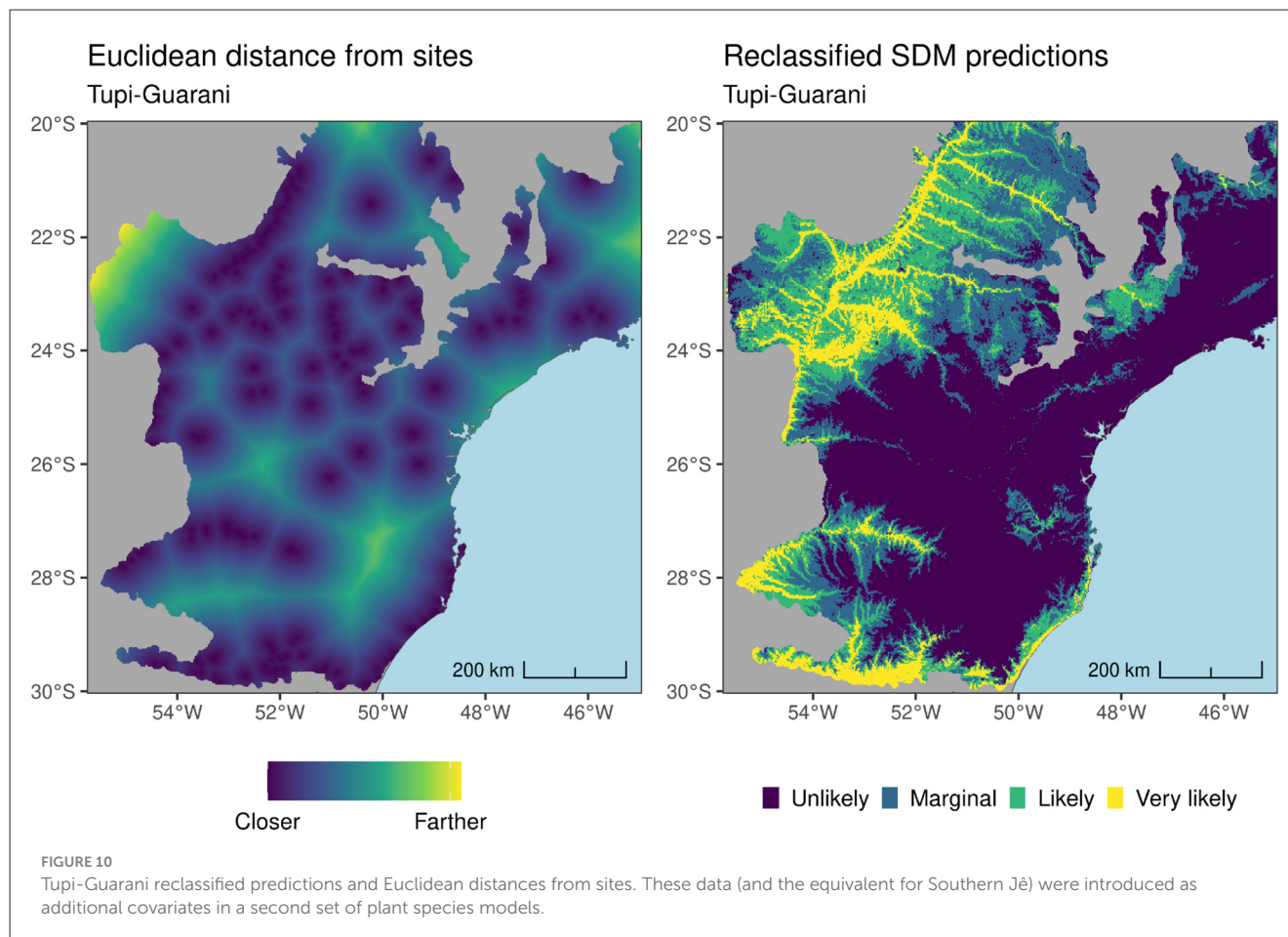
The impact of introducing ADCs on plant species model performance varied over different bootstrap runs (see below) but in some cases provided a considerable boost compared





to environment only models. The strength and nature of the relationship between ADCs and a single performance-boosted plant species model is shown in [Figure 11](#). In the case of *Araucaria*

angustifolia (top row), there is a clear and important (permutation importance of ~30%), positive relationship between reclassified Southern Jê predictions and predicted *Araucaria angustifolia*



likelihood of occurrence, which increases in step with predicted Southern Jê occurrence as it progresses from “Unlikely” to “Marginal” to “Likely,” before it plateaus in response to the highest value of “Very Likely.” A similar though less important relationship also exists between reclassified Southern Jê predictions and *Ilex paraguariensis* and *Palicourea sessilis* prediction responses. Positive relationships are observed between reclassified Tupi-Guarani predictions and *Tabernaemontana catharinensis* (~30% permutation importance) and reclassified Southern Jê predictions and *Tillandsia stricta* (~30% permutation importance). In the latter case, increasing reclassified Tupi-Guarani prediction values also translates into higher prediction scores. Overall, Euclidean distances from Southern Jê sites were almost universally unimportant across all the plant species models tested and as a result do not feature on [Figure 11](#). Euclidean distances from Tupi-Guarani sites feature as marginally important (< 10% permutation importance) across most plant species models. *Araucaria angustifolia* prediction values are high near to these sites and increase in strength up to a certain threshold (0.75) before they decrease rapidly, a pattern also shared with *Ilex paraguariensis* and *Palicourea sessilis*. The positive impact of ADCs on model P/E curves can also be observed and is discussed further in [Supplementary Data Sheet 1](#).

The above results indicate that the introduction of ADCs can frequently lead to substantial model improvement and therefore should be considered by analysts seeking to maximize

ecological model performance. However, introducing ADCs does not guarantee improvement for all species and in some bootstrap runs can be seen to degrade model performance across various evaluators. The variable benefit of introducing ADCs observed over different bootstrap runs likely occurs as a result of the weighted random sampling of background points used to train the model, the distribution of which are known to heavily impact Maxent model performance ([Barber et al., 2022](#); [Merow et al., 2013](#); [Valavi et al., 2022](#)). Observed improvements to model performance by the introduction of ADCs may be specific to one particular distribution of randomly sampled background points, making it difficult to ascertain whether ADCs will offer performance boosts more generally. To address this, two-sided Student’s *t*-tests were undertaken comparing all evaluation statistics from models, both with and without ADCs, across all bootstrapped runs. Of the 11 plant species models subjected to the phase two modeling routine, six showed a statistically significant difference across at least one of the performance evaluators measured (e.g., median CBI_{Test}). [Figure 12](#) shows the overall distribution of evaluators across all bootstrapped runs as well as the significance of the *t*-test’s *p* values for these six species. In terms of CBI~Test, statistically significant increases can be observed when introducing in reclassified Amerindian predictions as ADCs in models of *Araucaria angustifolia* (median, +0.01), *Myrcianthes pungens* (average, +0.04) and *Tillandsia stricta* (average, +0.001). For *Araucaria angustifolia* a statistically significant negative impact

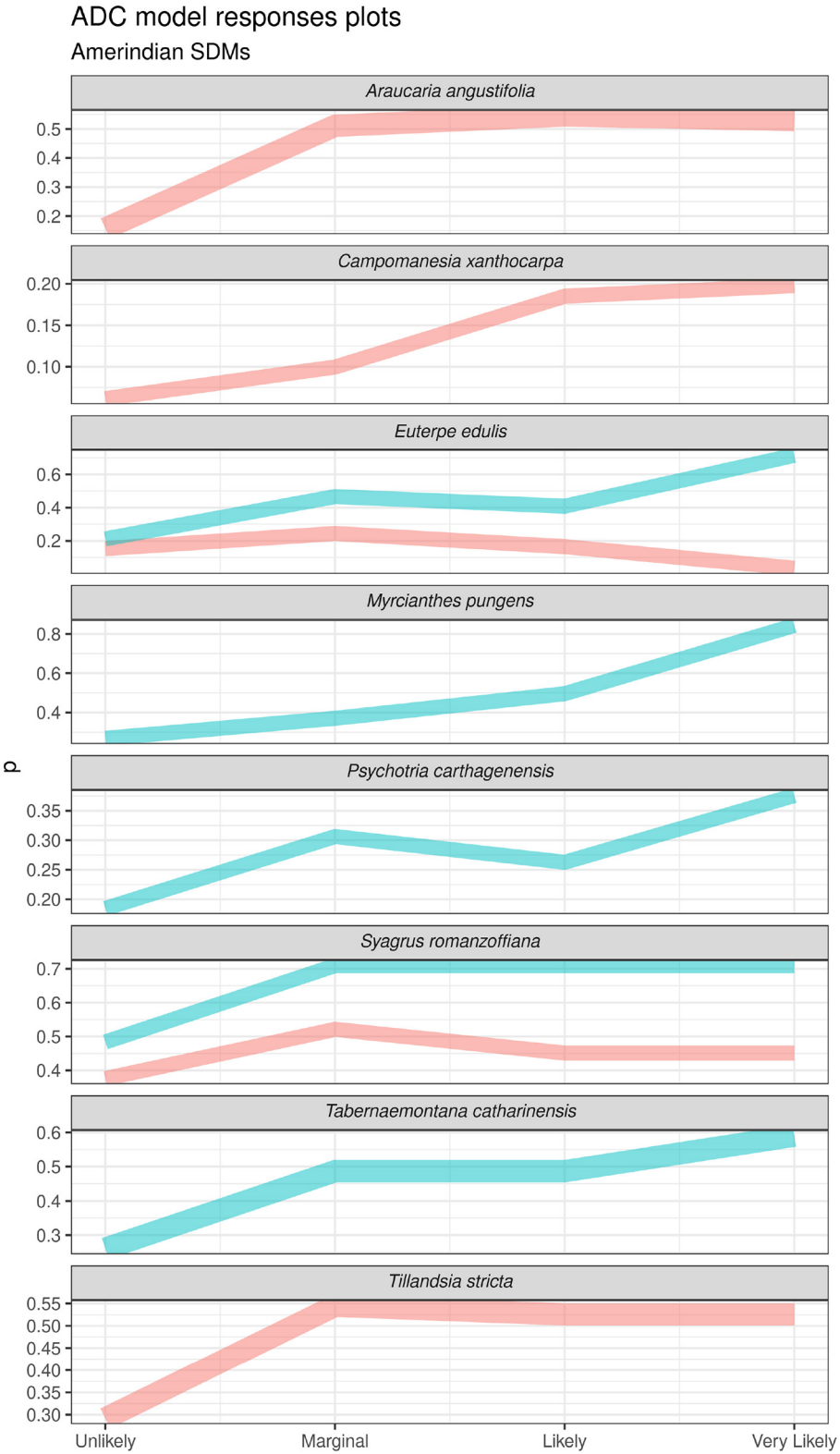


FIGURE 11
Response plots from plant species models showing model responses to ADCs. Responses are taken from the model with the highest permutation importance scores for ADC, with a minimum importance of 10%. Note, the ADC of euclidean distances from Amerindian sites did not acheive a permutation importance score of > 10% across any iterations and so no responses are shown.

on CBI_{Train} is also observed (though the size of this effect is practically negligible at -0.001). The standard deviation of CBI_{Test} for models of *Myrcianthes pungens* and *Tillandsia stricta* using reclassified Amerindian predictions were also degraded compared to environment only models, though the magnitude of this effect was only notable for the former ($+0.01$). This indicates that whilst ADCs increased the predictive performance of these models, they also reduced their overall stability—meaning they better predicted some spatial folds at the expense of others. Reclassified Amerindian predictions negatively impacted CBI_{Train} for *Araucaria angustifolia* (though the magnitude was marginal at -0.003), whilst they marginally increased CBI_{Train} alone for *Tabernaemontana catharinensis* ($+0.003$) and *Campomanesia xanthocarpa* ($+0.009$), indicating that introducing ADCs caused slight overfitting of these models.

5 Discussion

5.1 Amerindian predictions

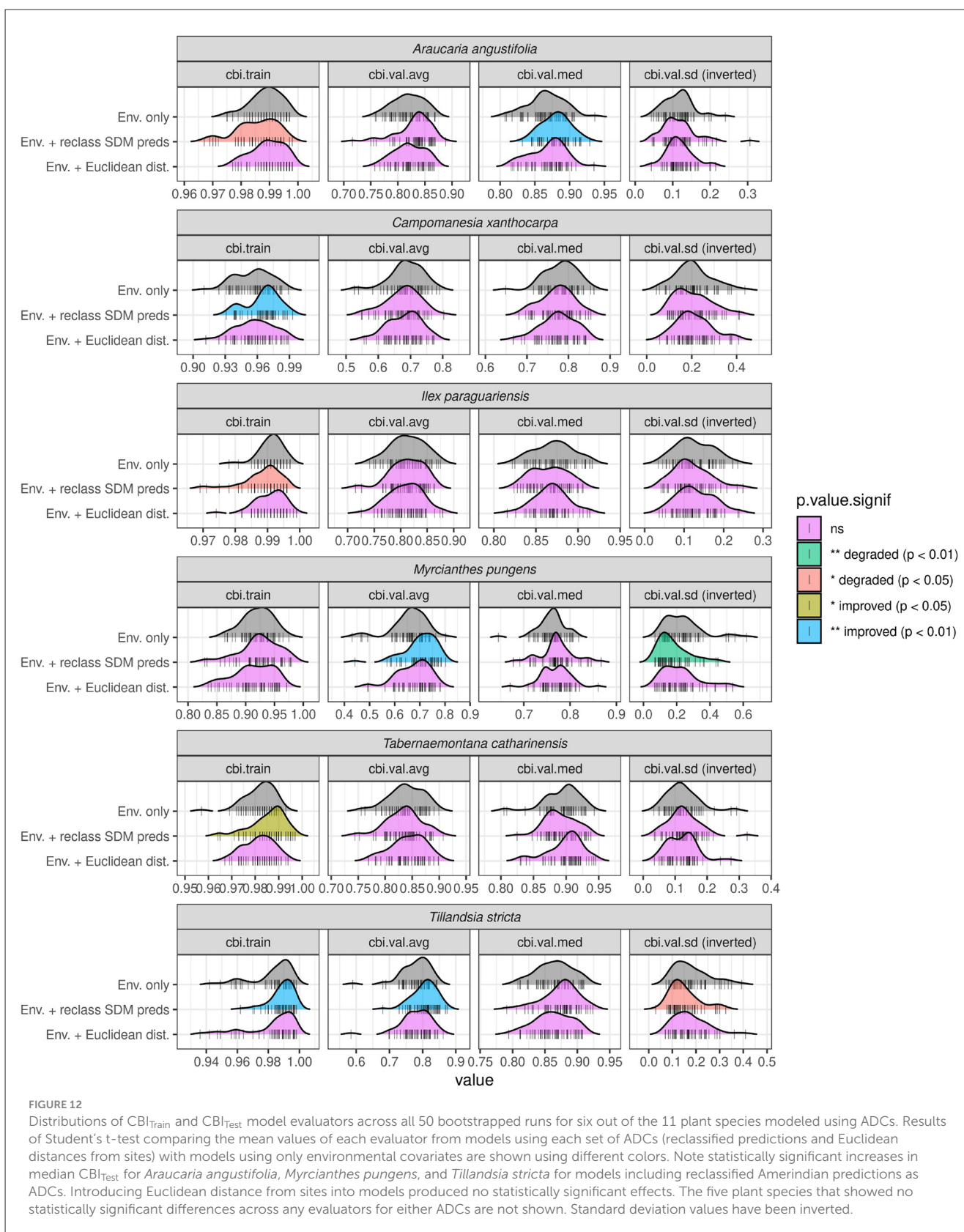
We preface our discussion by acknowledging that our models are based on partial presence data of pre-Colonial Indigenous people in southern Brazil. We made the affirmative methodological decision to make use of dated sites with known cultural affiliations, the assessment of which is based on expert judgement. Nevertheless, our study area encompasses regions, in particular the coastal strip between Florianópolis and São Paulo, as well as inland on the São Paulo plateau and in the Paraíba Valley, which had documented Tupi-Guarani presences in the sixteenth century (Fausto, 1992; Monteiro, 1992), but for which we lack presence data of the kind outlined above. The models—based as they are on partial and largely archaeological data that target pre-Columbian legacy effects—do not imply that there was no Guarani or Tupinambá history of occupation in these regions. Despite ambiguities in our dataset, we affirm that historical accounts indicate dense Colonial-era occupation along the coastal strip by Indigenous people. The following discussion is based on the areas for which we can analytically discern the clearest associations between pre-Columbian cultures and specific plants, and our interpretations are applicable to these areas alone except where explicitly stated. We anticipate that improvements in data coverage and quality will refine the propositions we tentatively advance here.

Insights about the importance of modeled environmental predictors—and the difference in terms of these between Tupi-Guarani and Southern Jê Amerindian groups—help to further characterize the kinds of environments favored by these traditions. Metrics from two commonly used evaluators (AUC and CBI) were compared against differing model resolutions and across plant species with a range of prevalence values. As other studies have demonstrated (Meynard et al., 2019), per species, median AUC_{Test} was seen to increase in line with coarsening resolution, further emphasizing the issues with this as an evaluator for presence-only models. Reclassification of continuous model predictions using the shape and CI intervals of the P/E curve is an effective way to visualize prediction data from presence-only models. As would be expected, niche overlap calculations indicate that plant species

with more established cultural or medical uses tend to share similar niches to either of the two Amerindian traditions modeled (see Supplementary Figure S1)—though further modeling of non-medicinal plant species is needed to establish the veracity of this relationship as well as the existence of possible causal, as opposed to correlative, links.

Most importantly, the results of this study indicate that the inclusion of archaeological data in ecological modeling frameworks can significantly increase model performance, both according to summary evaluation statistics such as CBI_{Test} and through examination of P/E curves. More widely, this research contributes methodologically to the wider emergent subdiscipline of archaeoecology (Crabtree and Dunne, 2022). Our results indicate that reclassified predictions of Amerindian likelihood of occurrence improved model performance more than using distances from known Amerindian sites alone. Response plots from *Araucaria angustifolia* models showed increased prediction strength in step with high and/or very high likelihood of occurrence for the Southern Jê. Though the impact of ADCs varied according to the distribution of the background points used to train Maxent models, on average, ADCs significantly increased the median CBI_{Test} value of models for a number of plant species of documented indigenous value, a reliable indicator of the predictive ability of a presence-only model on unseen data. The performance improvement and relative importance of reclassified Amerindian SDMs across all model covariates compared to simple Euclidean distances to Amerindian sites (see Figure 11 above) accords with previously published archaeoecological studies utilizing a similar a methodology (Tulowiecki and Larsen, 2015; Tulowiecki et al., 2022). The disparity is likely due to the reclassified SDM predictions better capturing the geographical extent of zones of probable Amerindian environmental impact, including at locales containing as yet undetected Amerindian sites. The paucity of confirmed archaeological sites at locations that appear to show environmental impacts elsewhere associated with proximity to archaeological sites was highlighted by Levis et al. (2017): 927 as a probable contributor to the relatively modest explanatory power of models based on Euclidean distances from archaeological sites in the Amazon basin. Models that incorporate (reclassified) SDM predictions, or other non-Euclidean distance based predictors [e.g. “accessibility” maps based on least-cost paths from archaeological sites, see Tulowiecki and Larsen (2015)], could usefully be applied in such regions to better characterize the impact of archaeological land management practices on modern day vegetation composition.

Notwithstanding the apparent differences in environmental preference between Southern Jê settlement sites and *Araucaria angustifolia* habitat (see Figure 7), improvements to the latter model following the introduction of ADCs may reflect the fact that particular archaeological traditions and particular plant species shared (and continue to share) similar ecological niches, with the ADCs providing enough useful information to the model to enable it to increase/decrease predictions at locations that contain/do not contain species presences but are otherwise difficult to account for using the remaining environmental covariates. An obvious next step to confirm these findings would involve diversifying the modeling methods deployed here (to include e.g. random forests, SVM, GAM, GLM etc.) and/or working with more systematically



collected plant species occurrence and ideally absence data (see Tulowiecki and Larsen, 2015). Further experimentation with background point generation would also be advisable, as a lack

of background points around the northernmost Tupi-Guarani presences available—due to a relative dearth of unattributed sites here—likely meant the available environmental conditions in this

area were not adequately sampled (see Figure 4). Model prediction data are freely available to facilitate further research into Pre-Columbian human-environmental interactions in the southern Atlantic Forest (<https://doi.org/10.5281/zenodo.14955744>). High likelihood zones within Amerindian model predictions may also usefully warrant further archaeological survey or desk-based assessments, where previously published studies adopting a similar approach have located hitherto undetected archaeological sites via examination of available LiDAR (e.g., Walker et al., 2023).

5.2 Implications

Our use of Amerindian-derived covariates (ADCs) underscores some important features of pre-Columbian legacy effects in the southern Atlantic Forest. While improvements were detected among several culturally important species (see Supplementary Table S1 and Supplementary Figure S2), ADCs consistently enhanced model predictive power for three species: *Araucaria angustifolia*, *Tillandsia stricta* and *Myrcianthes pungens* (Figure 12). First, our models' detection of significant niche overlap between pre-Columbian southern Jê sites and *Araucaria* accords well with previous archaeological and palaeoecological findings in the study region (Behling, 2002; Bitencourt, 2006; Iriarte and Behling, 2007; Robinson et al., 2018; Lauterjung et al., 2018; Azevedo and Scheel-Ybert, 2020; Pereira Cruz et al., 2020). Beyond the evident dietary importance of *Araucaria* seeds to these groups (Métraux, 1946; Beber, 2004; Schmitz et al., 2013; Iriarte et al., 2017a), *Araucaria* also played an important role in the wider economy through provisioning of firewood and construction materials (Azevedo and Scheel-Ybert, 2020). Ethnohistoric sources note that among the Kaingang, access to these coveted resources were centrally controlled by community leaders (Mabilde, 1988). While the direction of causality remains challenging to infer (i.e., did Late Holocene *Araucaria* forest expansion encourage highland population growth, or did expanding populations promote forest growth to enhance food supply?), the fact that the relationship between these phenomena are salient features of our models indicates that our inferences are robust. No improvement in model performance upon incorporating ADCs might demonstrate the limitations of our presence-only based modeling approach and/or gaps within the underlying training data, or even call into question long-held assumptions about the socio-ecology of southern Brazil.

Second, the relationship between the epiphytic *Tillandsia stricta* (Family: Bromeliaceae) and human presence is illustrative of the potential contribution of macroecological modeling against the backdrop of increasing focus on and reclamation of traditional ecological knowledge (de Andrade et al., 2021; Klein et al., 2022; Pavão et al., 2021). In general, this genus has numerous medicinal uses among Guarani groups of southern Brazil, including as a topical cosmetic, anti-inflammatory, and diuretic agent (Estrella-Parra et al., 2019). Bromeliads, including *Tillandsia* spp., are frequently traded as ornamentals and may have commercial value to traditional communities as part of sustainable forest management practices (Peralta-Kulik et al., 2023). Although they are recognized for their cultural importance, Bromeliaceae in general are rarely, if ever, reported in regional palaeoecological records, possibly due to comprising a very small percentage

of overall pollen rain. In contrast to the imposing and iconic *Araucaria* tree, a keystone species of the southern Atlantic Forest, *Tillandsia* spp. are relatively unobtrusive. The nutritious fruit of *Myrcianthes pungens* is well-documented as a food source for both contemporary Indigenous groups in east Paraguay (Edeb Piragi, 2011) and historically amongst the Guarani in southern Brazil (Pereira et al., 2016), and the species is therefore an obvious candidate for propagation and dispersal by Indigenous populations in the region within extant and past agroforestry systems. In summary, our use of modern plant inventories cross-referenced to Indigenous Amerindian taxonomic knowledge enabled an inclusive approach that we argue complements other sources of information on human-environmental interaction. It has generated models with enhanced dimensionality and scope that incorporated best practices in species distribution modeling methodologies. In the process, we have improved our knowledge of potential pre-Columbian legacy effects on contemporary forest composition. Our results argue the case for ecological modelers to more routinely consider including archaeological data within contemporary floristic SDMs, especially in landscapes with a long-standing human presence. Robust predictions, incorporating human presence as an important component of tropical ecosystems (Montoya et al., 2020; Scerri et al., 2022), may also be used to facilitate or strengthen further downstream analyses involving recent or contemporary floristic inventories to characterize pre-Columbian environmental legacy effects (e.g., Levis et al., 2017), with important implications for conservation and policy-making more broadly. As biodiversity outcomes are enhanced by the promotion of Indigenous land use strategies (Benzeev et al., 2023) the potential of palaeodata to contribute or even suggest alternatives to modern land use patterns is clear (Silva et al., 2022).

Data availability statement

The datasets presented in this article are not readily available because data collection, by a third party, is in progress at the time of publication. Requests to access the datasets should be directed to the MAPHSA project (Universitat Pompeu Fabra, Barcelona, Spain).

Author contributions

BH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. HB: Writing – review & editing. JG: Writing – review & editing. AR: Writing – review & editing. PRO: Writing – original draft, Writing – review & editing. PRi: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The research was funded by the UK Arts and Humanities Research Council (AH/X002217/1).

Acknowledgments

Freg Stokes provided invaluable constructive criticism on the shortcomings of the training data used for Amerindian models. In particular, he highlighted that the absence of coastal pre-Columbian Tupi-Guarani sites was an artifact of the data, as many sites are confirmed to have existed within the historical record.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AC declared a past co-authorship with the author PPro to the handling editor.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

References

- Akinwande, M. O., Dikko, H. G., and Samson, A. (2015). Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. *Open J. Statist.* 5:754. doi: 10.4236/ojs.2015.57075
- Allouche, O., Steinitz, O., Rotem, D., Rosenfeld, A., and Kadmon, R. (2008). Incorporating distance constraints into species distribution models. *J. Appl. Ecol.* 45, 599–609. doi: 10.1111/j.1365-2664.2007.01445.x
- Amatulli, G., Domisch, S., Tuanmu, M. N., Parmentier, B., Ranipeta, A., Malczyk, J., et al. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci. Data* 5:180040. doi: 10.1038/sdata.2018.40
- Arroyo-Kalin, M. (2017). Amazonian Dark Earths, in *Archaeological Soil and Sediment Micromorphology*, (John Wiley & Sons, Ltd). 345–57. doi: 10.1002/9781118941065.ch33
- Azevedo, L. W., and Scheel-Ybert, R. (2020). Contributions to Proto-Jê archaeology in the southern Brazilian highlands: wood, fire, and landscape. *Latin Am. Antiq.* 31, 325–41. doi: 10.1017/laq.2020.1
- Baker, D. J., Maclean, I. M. D., Goodall, M., and Gaston, K. J. (2022). Correlations between spatial sampling biases and environmental niches affect species distribution models. *Glob. Ecol. Bio.* 31, 1038–50. doi: 10.1111/geb.13491
- Balée, W. (2013). *Cultural Forests of the Amazon: A Historical Ecology of People and Their Landscapes*. Tuscaloosa, AL: University of Alabama Press. doi: 10.2307/jj.30347164
- Barber, R. A., Ball, S. G., Morris, R. K. A., and Gilbert, F. (2022). Target-group backgrounds prove effective at correcting sampling bias in maxent models. *Divers. Distr.* 28, 128–41. doi: 10.1111/ddi.13442
- Barbet-Massin, M., Jiguet, F., Hélène Albert, C., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Met. Ecol. Evol.* 3, 327–38. doi: 10.1111/j.2041-210X.2011.00172.x
- Barlow, J., Gardner, T. A., Lees, A. C., Parry, L., and Peres, C. A. (2012). How pristine are tropical forests? an ecological perspective on the pre-columbian human footprint in Amazonia and implications for contemporary conservation. *Biol. Conserv.* 151, 45–49. doi: 10.1016/j.biocon.2011.10.013
- Beber, M. V. (2004). *O Sistema De Assentamento Dos Grupos Ceramistas Do Planalto Sul-Brasileiro O Caso Da Tradição Taquara/Itararé*. PhD thesis, São Leopoldo: Universidade do Vale do Rio dos Sinos.
- Behling, H. (2002). South and Southeast Brazilian Grasslands during late quaternary times: a synthesis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 177, 19–27. doi: 10.1016/S0031-0182(01)00349-2
- Behling, H., DePatta Pillar, V., and Girardi Bauermann, S. (2005). Late quaternary grassland (campos), gallery forest, fire and climate dynamics, studied by pollen, charcoal and multivariate analysis of the São Francisco de Assis core in western Rio Grande do Sul (Southern Brazil). *Rev. Palaeobot. Palynol.* 133, 235–48. doi: 10.1016/j.revpalbo.2004.10.004
- Behling, H., DePatta Pillar, V., Orlóci, L., and Bauermann, G. S. (2004). Late quaternary araucaria forest, grassland (campos), fire and climate dynamics, studied by high-resolution pollen, charcoal and multivariate analysis of the Cambará Do Sul Core in Southern Brazil. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 203, 277–97. doi: 10.1016/S0031-0182(03)00687-4
- Benzeer, R., Zhang, S., Artur Rauber, M., Vance, E. A., and Newton, P. (2023). formalizing tenure of indigenous lands improved forest outcomes in the Atlantic Forest of Brazil. *PNAS Nexus* 2:pgac287. doi: 10.1093/pnasnexus/pgac287
- Bitencourt, A. L. (2006). possible prehistoric anthropogenic effect on araucaria angustifolia (bert.) O. Kuntze expansion during the late holocene. *Revista Brasileira de Paleontologia* 9, 109–16. doi: 10.4072/rbp.2006.1.12
- Bonomo, M., Angrizani, R. C., Apolinare, E., and Noelli, F. S. (2015). A model for the guaraní expansion in the La Plata Basin and Littoral zone of Southern Brazil. *Quat. Intern.* 356, 54–73. doi: 10.1016/j.quaint.2014.10.050
- Boria, R. A., Olson, L. E., Goodman, S. M., and Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological Niche Models. *Ecol. Model.* 275, 73–77. doi: 10.1016/j.ecolmodel.2013.12.012
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecol. Model.* 157, 281–300. doi: 10.1016/S0304-3800(02)00200-4
- Brochado, J. J. P. (1984). *An Ecological Model of the Spread of Pottery and Agriculture into Eastern South America*. PhD thesis, United States – Illinois: University of Illinois at Urbana-Champaign.
- Bucklin, D. N., Basille, M., Benschoter, A. M., Brandt, L. A., Mazzotti, F. J., Romañach, S. S., et al. (2015). Comparing species distribution models constructed with different subsets of environmental predictors. *Divers. Distr.* 21, 23–35. doi: 10.1111/ddi.12247
- Bueno, N. R., Oliveira Castilho, R., Da Costa, R. B., Pott, A., Pott, V. J., Scheidt, G. N., et al. (2005). Medicinal plants used by the kaiowá and guarani indigenous populations in the Caarapó Reserve, Mato Grosso Do Sul, Brazil. *Acta Botanica Brasílica* 19, 39–44. doi: 10.1590/S0102-33062005000100005
- Capriles, J. M., García, M., Valenzuela, D., Domic, A. I., Kistler, L., Rothhammer, F., et al. (2022). Pre-Columbian Cultivation of vegetatively propagated and fruit tree tropical crops in the Atacama Desert. *Front. Ecol. Evol.* 10:993630. doi: 10.3389/fevo.2022.993630
- Cardoso, J. M., Merencio, F., Villagran, X., Wesolowski, V., Estevam, R., Fuller, B. T., et al. (2024). Late Shellmound occupation in southern Brazil: a

multi-proxy study of the galheta iv archaeological site. *PLOS ONE* 19:e0300684. doi: 10.1371/journal.pone.0300684

Carlucci, M. B., Marcilio-Silva, V., and Marcelo Torezan, J. (2021). "The Southern Atlantic Forest: Use, Degradation, and Perspectives for Conservation," In *The Atlantic Forest: History, Biodiversity, Threats and Opportunities of the Mega-diverse Forest*, eds. Marcia C. M. Marques and Carlos E. V. Grelle, (Cham: Springer International Publishing) 91–111. doi: 10.1007/978-3-030-55322-7_5

Cassino, M. F., Shock, M. P., Furquim, L. P., Ortega, D. D., Machado, J. S., Madella, M., and Clement, C. R. (2021). "Archaeobotany of Brazilian Indigenous Peoples and Their Food Plants," In *Local Food Plants of Brazil*, eds. M. C. Medeiros Jacob and U. P. Albuquerque, (Cham: Springer International Publishing) 127–59. doi: 10.1007/978-3-030-69139-4_8

Clement, C. R., Denevan, W. M. D., Heckenberger, M. J., Junqueira, A. B., Neves, E. G., Teixeira, W. G. et al. (2015). The domestication of amazonia before European Conquest. *Proceed. Royal Soc. B. Biol. Sci.* 282. doi: 10.1098/rspb.2015.0813

Corteletti, R. (2013). *Projeto Arqueológico Alto Canoas-Paraca: Um Estudo Da Presença Jê No Planalto Catarinense*. São Paulo: Universidade de São Paulo.

Corteletti, R., Dickau, R., DeBlasis, P., and Iriarte, J. (2015). Revisiting the economy and mobility of southern Proto-Jê (Taquara-Itararé) groups in the Southern Brazilian Highlands: starch grain and phytoliths analyses from the Bonin Site, Urubici, Brazil. *J. Archaeol. Sci.* 58, 46–61. doi: 10.1016/j.jas.2015.03.017

Crabtree, S. A., and Dunne, J. A. (2022). Towards a science of archaeoecology. *Trends Ecol. Evol.* 37, 976–84. doi: 10.1016/j.tree.2022.07.010

de Andrade, J. H. C., Rodrigues, J., Benites, A., Benites, C., Acosta, A., Benites, M., et al. (2021). Notes on Current Mbyá-Guarani medicinal plant exchanges in Southern Brazil. *J. Ethnobiol. Ethnomed.* 17:38. doi: 10.1186/s13002-021-00465-w

De Souza, J. G., and Riris, P. (2021). Delayed demographic transition following the adoption of cultivated plants in the eastern la plata basin and Atlantic Coast, South America. *J. Archaeol. Sci.* 125:105293. doi: 10.1016/j.jas.2020.105293

Demján, P., Dreslerová, D., Kolár, J., Chuman, T., Romportl, D., Trnka, M., and Lieskovský, T. (2022). Long time-series ecological niche modelling using archaeological settlement data: tracing the origins of present-day landscape. *Appl. Geogr.* 141:102669. doi: 10.1016/j.apgeog.2022.102669

Dias, A. S. (2012). Hunter-gatherer occupation of south brazilian atlantic forest: paleoenvironment and archaeology. *Quaternary Int.* 256, 12–18. doi: 10.1016/j.quaint.2011.08.024

Edeb Piragi, P. (2011). Dynamiques identitaires anciennes et actuelles chez les Aché du Paraguay oriental : éléments de compréhension. *J. de la Société des américanistes* 97, 231–85. doi: 10.4000/jsa.11961

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–51. doi: 10.1111/j.2006.0906-7590.04596.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. N., and Yates, C. J. (2011). A statistical explanation of maxent for ecologists. *Diver. Distr.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Erickson, C. L. (2000). An artificial landscape-scale fishery in the bolivian Amazon. *Nature* 408, 190–93. doi: 10.1038/35041555

Estrella-Parra, E., Flores-Truz, M., Blancas-Flores, G., Koch, S. D., and Alarcón-Aguilar, F. G. (2019). The cillandria genus: history, uses, chemistry, and biological activity. *Bol Latinoam Caribe Plant Med Aromat* 18, 239–64.

Fausto, C. (1992). *Fragmentos de história e cultura tupinambá: da etnologia como instrumento crítico de conhecimento etno-histórico*. In *História dos índios no Brasil, edited by Núcleo de História Indígena e do Indigenismo*, 381–96. Cunha, Manuela Carneiro da (org.). São Paulo: Companhia das Letras, Secretaria Municipal de Cultura, FAPESP.

Feng, X., Park, D. S., Liang, Y., Pandey, R., and Papeş, M. (2019). Collinearity in ecological niche modeling: confusions and challenges. *Ecol. Evol.* 9, 10365–76. doi: 10.1002/ece3.5555

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: New 1-Km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–15. doi: 10.1002/joc.5086

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuisen, H. T., Verelst, L., and Wiberg, D. (2008). Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy 10.

Fithian, W., and Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *Ann. Appl. Statist.* 7, 1917–39. doi: 10.1214/13-AOAS667

Flores, B. M., Montoya, E., Sakschewski, B., Nascimento, N., Staal, A., Betts, R. A., et al. (2024). Critical transitions in the amazon forest system. *Nature* 626, 555–64. doi: 10.1038/s41586-023-06970-0

GBIF.org (2024). GBIF Occurrence Download.

Gessert, S., Iriarte, J., Carlos Ríos, R., and Behling, H. (2011). Late holocene vegetation and environmental dynamics of the araucaria forest region in misiones province, ne argentina. *Rev. Palaeobot. Palynol.* 166, 29–37. doi: 10.1016/j.revpalbo.2011.04.006

Gillson, L., and Marchant, R. (2014). From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology. *Trends Ecol. Evol.* 29, 317–25. doi: 10.1016/j.tree.2014.03.010

Grimmett, L., Whitsed, R., and Horta, A. (2020). Presence-only species distribution models are sensitive to sample prevalence: evaluating models using spatial prediction stability and accuracy metrics. *Ecol. Model.* 431:109194. doi: 10.1016/j.ecolmodel.2020.109194

Hirzel, A. H., Lay, L. L., Helfer, V., Randin, C., and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–52. doi: 10.1016/j.ecolmodel.2006.05.017

Howey, M. C. L., Palace, M. W., and McMichael, C. H. (2016). Geospatial modeling approach to monument construction using Michigan from A.D. 1000–1600 as a case study. *Proceed. Nat. Acad. Sci.* 113, 7443–48. doi: 10.1073/pnas.1603450113

Iriarte, J., and Behling, H. (2007). The expansion of araucaria forest in the southern brazilian highlands during the last 4000 years and its implications for the development of the taquara/itararé tradition. *Environ. Archaeol.* 12, 115–27. doi: 10.1179/174963107x226390

Iriarte, J., Christopher Gillam, J., and Marozzi, O. (2008). Monumental burials and memorial feasting: an example from the southern brazilian highlands. *Antiquity* 82, 947–61. doi: 10.1017/S0003598X00097702

Iriarte, J., DeBlasis, J., De Souza, J. G., and Corteletti, R. (2017a). Emergent complexity, changing landscapes, and spheres of interaction in Southeastern South America during the Middle and Late Holocene. *J. Archaeol. Res.* 25, 251–313. doi: 10.1007/s10814-016-9100-0

Iriarte, J., Elliott, S., Maezumi, S. Y., Alves, D., Gonda, R., Robinson, M., et al. (2020). The origins of amazonian landscapes: plant cultivation, domestication and the spread of food production in tropical South America. *Quat. Sci. Rev.* 248:106582. doi: 10.1016/j.quascirev.2020.106582

Iriarte, J., Moehlecke Copé, S., Fradley, M., Lockhart, J. J., and Gillam, J. C. (2013). Sacred landscapes of the Southern Brazilian highlands: understanding southern proto-jê mound and enclosure complexes. *J. Anthropol. Archaeol.* 32, 74–96. doi: 10.1016/j.jaa.2012.10.003

Iriarte, J., Smith, R. J., de Souza, J. G., Mayle, F. E., Whitney, B. S., Cárdenas, M. L. (2017b). Out of Amazonia: late-holocene climate change and the tupi-guarani trans-continental expansion. *Holocene* 27, 967–75. doi: 10.1177/0959683616678461

Jeske-Pieruschka, V., and Behling, H. (2012). Palaeoenvironmental history of the são francisco de paula region in southern brazil during the late quaternary inferred from the rincão das cabritas core. *Holocene* 22, 1251–62. doi: 10.1177/0959683611414930

Jiménez, L., and Soberón, J. (2020). Leaving the area under the receiving operating characteristic curve behind: an evaluation method for species distribution modelling applications based on presence-only data. *Met. Ecol. Evol.* 11, 1571–86. doi: 10.1111/2041-210X.13479

Jiménez-Valverde, A. (2012). Insights into the Area Under the Receiver Operating Characteristic Curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 498–507. doi: 10.1111/j.1466-8238.2011.00683.x

Johnson, J. M., Munasinghe, D., Eyelade, D., and Cohen, S. (2019). An integrated evaluation of the national water model (NWM)–height above nearest drainage (HAND) flood mapping methodology. *Nat. Hazards Earth Syst. Sci.* 19, 2405–20. doi: 10.5194/nhess-19-2405-2019

Joint Research Centre (2022). *Phenology Dataset*. GEO Knowledge Hub.

Júnior, P. D. M., and Nóbrega, C. C. (2018). Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE* 13:e0202403. doi: 10.1371/journal.pone.0202403

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., et al. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Met. Ecol. Evol.* 12, 1602–8. doi: 10.1111/2041-210X.13628

Keller, H. A., Clifford, V., and Araujo, J. J. (2010). The diversity and sustainability of ornamental plants traded by guarani communities of Misiones province, Argentina.

Klein, T., Shiratori, K., and Amaral, T. (2022). Yvoty: flores e outras plantas do povo Guarani Mbya. *Maloca: Revista de Estudos Indígenas* 5:e022011. doi: 10.20396/maloca.v5i00.15503

Kramer-Schadt, S., Niedballa, S., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., et al. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distr.* 19, 1366–79. doi: 10.1111/ddi.12096

Lauterjung, M. B., Bernardi, A. P., Montagna, T., Candido-Ribeiro, R., Freitas da Costa, N. C., Mantovani, A., et al. (2018). Phylogeography of Brazilian Pine (*Araucaria Angustifolia*): Integrative Evidence for Pre-Columbian Anthropogenic Dispersal. *Tree Genet. Genom.* 14:36. doi: 10.1007/s11295-018-1250-4

Ledru, M.-P., Montade, V., Blanchard G., and Hély, C. (2016). Long-term spatial changes in the distribution of the Brazilian Atlantic Forest. *Biotropica* 48, 159–69. doi: 10.1111/btp.12266

- Lehner, B., and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Proces.* 27, 2171–86. doi: 10.1002/hyp.9740
- Leroy, B. (2023). Choosing presence-only species distribution models. *J. Biogeogr.* 50, 247–50. doi: 10.1111/jbi.14505
- Levis, C., Costa, F. R. C., Bongers, F., Peña-Claros, M., Clement, C. R., Junqueira, A. B., et al. (2017). Persistent effects of pre-columbian plant domestication on amazonian forest composition. *Science* 355, 925–31. doi: 10.1126/science.aal0157
- Li, W., and Guo, Q. (2013). How to assess the prediction accuracy of species presence-absence models without absence data? *Ecography* 36, 788–99. doi: 10.1111/j.1600-0587.2013.07585.x
- Lins, J., Lima, H. P., Baccaro, F. B., Kinupp, V. F., Shepard, G. H. Jr, and Clement, C. R. (2015). Pre-Columbian floristic legacies in modern homegardens of central amazonia. *PLoS ONE* 10:e0127067. doi: 10.1371/journal.pone.0127067
- Liu, C., White, M., and Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. edited by richard pearson. *J. Biogeogr.* 40, 778–89. doi: 10.1111/jbi.12058
- Lobo, J. M., Jiménez-Valverde, A., and Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–51. doi: 10.1111/j.1466-8238.2007.00358.x
- Lombardo, U., Arroyo-Kalin, M., Schmidt, M., Huisman, H., Lima, H. P., Moraes, C. L., et al. (2022). Evidence confirms an anthropic origin of amazonian dark earths. *Nat. Commun.* 13:3444.
- Lombardo, U., and Prümers, H. (2010). Pre-Columbian human occupation patterns in the eastern plains of the llanos de moxos, bolivian amazonia. *J. Archaeol. Sci.* 37, 1875–85. doi: 10.1016/j.jas.2010.02.011
- Mabilde, A. (1988). O Índio Kaingang Do Rio Grande Do Sul No Final Do século XIX. *Documentos* 2, 141–72.
- Maizumi, S. Y., Alves, D., Robinson, M., Gregorio de Souza, J., Levis, C., Barnett, R. L., et al. (2018). The Legacy of 4,500 years of polyculture agroforestry in the Eastern Amazon. *Nat. Plants* 4, 540–47. doi: 10.1038/s41477-018-0205-y
- Malhi, Y., Gardner, T. A., Goldsmith, G. R., Silman, M. R., and Zelazowski, P. (2014). Tropical forests in the anthropocene. *Ann. Rev. Environ. Res.* 39, 125–59. doi: 10.1146/annurev-environ-030713-155141
- Manzoor, S. A., Griffiths, G., and Lukac, M. (2018). Species distribution model transferability and model grain size – finer may not always be better. *Sci. Rep.* 8. doi: 10.1038/s41598-018-25437-1
- Marchioro, C. A., Santos, K. L., and Siminski, A. (2020). Present and future of the critically endangered *araucaria angustifolia* due to climate change and habitat loss. *Forest. Int. J. Forest Res.* 93, 401–10. doi: 10.1093/forestry/cpz066
- Marques, M. C. M., Trindade, W., Bohn, A., and Grelle, C. E. V. (2021). “The Atlantic Forest: An Introduction to the Megadiverse Forest of South America,” In *The Atlantic Forest: History, Biodiversity, Threats and Opportunities of the Megadiverse Forest*, eds. Marcia C. M. Marques and Carlos E. V. Grelle, (Cham: Springer International Publishing) 3–23. doi: 10.1007/978-3-030-55322-7_1
- Martínez Crovetto, R. N. (2012). Estudios Etnobotánicos V. Nombres De Plantas Y Su Utilidad Según Los Mbya Guaraní De Misiones, Argentina. *Bonplandia* 21, 109–33. doi: 10.30972/bon.2121282
- McDowell, R. W., Noble, A., Pletnyakov, P., and Haygarth, P. M. (2023). A global database of soil plant available phosphorus. *Sci. Data* 10:125. doi: 10.1038/s41597-023-02022-4
- McMichael, C. H., Palace, M. W., Bush, M. B., Braswell, B., Hagen, S., Neves, E. G., et al. (2014). Predicting Pre-Columbian anthropogenic soils in amazonia. *Proceed. Royal Soc. B. Biol. Sci.* 281:20132475. doi: 10.1098/rspb.2013.2475
- Merow, C., Smith, M. J., and Silander, J. A. Jr. (2013). A practical guide to maxent for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–69. doi: 10.1111/j.1600-0587.2013.07872.x
- Métraux, A. (1946). “The Caingang,” in *Handbook of South American Indians*, eds. J. H. Steward, (Washington, DC: Smithsonian Institution) 445–77
- Meynard, C. N., Leroy, B., and Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography* 42, 2021–36. doi: 10.1111/ecog.04385
- Milheira, R. G., and DeBlasis, P. (2014). “Tupi-Guarani Archaeology in Brazil,” in *Encyclopedia of Global Archaeology*, eds. C. Smith, (New York, NY: Springer) 7384–89. doi: 10.1007/978-1-4419-0465-2_2487
- Monteiro, J. M. (1992). “Os Guaraní e a história do Brasil meridional: séculos XVI-XVII,” in *História dos índios no Brasil*, eds. Núcleo de História Indígena e do Indigenismo, Cunha, Manuela Carneiro da (org.). São Paulo: Companhia das Letras, Secretaria Municipal de Cultura, FAPESP. 475–98.
- Montoya, E., Lombardo, U., Levis, C., Aymard, G. A., and Mayle, F. E. (2020). “Human Contribution to Amazonian Plant Diversity: Legacy of Pre-Columbian Land Use in Modern Plant Communities,” In *Neotropical Diversification: Patterns and Processes*, eds. Valenti Rull and Ana Carolina Carnaval, (Cham: Springer International Publishing) 495–520. doi: 10.1007/978-3-030-31167-4_19
- Morales, N. S., Fernández, I. C., and Baca-González, V. (2017). MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? a systematic review. *PeerJ* 5:e3093. doi: 10.7717/peerj.3093
- Nascimento, M. N., Heijink, B. M., Bush, M. B., Gosling, W. D., and McMichael, C. N. H. (2022). Early to mid-holocene human activity exerted gradual influences on amazonian forest vegetation. *Phil. Trans. Royal Soc. B. Biol. Sci.* 377:20200498. doi: 10.1098/rstb.2020.0498
- Nascimento, M. N., Teye, F. N. A., and McMichael, C. N. H. (2024). Indigenous and colonial influences on amazonian forests. *Plants People Planet* 6, 803–23. doi: 10.1002/ppp3.10515
- Noelli, F. S. (1993). *Sem Tekohá não há Tekó (Em Busca de Um Modelo Etnoarqueológico Da Subsistência e Da Aldeia Guarani Aplicado a Uma Área de Domínio No Delta Do Jacui-RS)*. PhD thesis, Porto Alegre: PUCRS.
- Noelli, F. S. (1998). The tupi: explaining origin and expansions in terms of archaeology and of historical linguistics. *Antiquity* 72, 648–63. doi: 10.1017/S0003598X00087068
- Noelli, F. S. (2004). La distribución geográfica de las evidencias arqueológicas Guaraní. *Revista de Indias* 64, 17–34. doi: 10.3989/revindias.2004.i230.408
- Noelli, F. S. (2005). “Rethinking Stereotypes and the History of Research on Jê Populations in South Brazil: An Interdisciplinary Point of View,” In *Global Archaeological Theory: Contextual Voices and Contemporary Thoughts*, eds. P. P. Funari, A. Zaranekin, and E. Stovel, (Boston, MA: Springer US) 167–90. doi: 10.1007/0-306-48652-0_12
- Noelli, F. S. (2008). The Tupi Expansion. In *The Handbook of South American Archaeology*, edited by Helaine Silverman and William H. Isbell, 659–70. New York, NY: Springer. doi: 10.1007/978-0-387-74907-5_33
- Park, D. S., Xie, Y., Thammavong, H. T., Tulaiha, R., and Feng, X. (2023). Artificial Hotspot Occurrence Inventory (AHOI). *J. Biogeogr.* 50, 441–49. doi: 10.1111/jbi.14543
- Pavão, S., Lopes, I. G., Vilharva, K. V., Peralta, A. P., da Silva Pedro, M., Benites, E., and Gislotti, L. J. (2021). Flora Medicinal Guarani E Kaiowá: Conhecimento Tradicional Como Forma De Resistência. *Espaço Ameríndio* 15, 160–60. doi: 10.22456/1982-6524.105223
- Peralta-Kulik, N., Amarilla Rodríguez, S. M., Pérez de Molas, L., Villalba, J. G., Peralta-Kulik, N., Amarilla Rodríguez, S. M., et al. (2023). Approaches to the economic valuation of non-timber products from the Alto Paraná Atlantic Forests, Paraguay. *Revista Chapingo Serie Ciencias Forestales y Del Ambiente* 29, 61–76. doi: 10.5154/r.chscfa.2022.12.085
- Pereira Cruz, A., Hettwer Giehl, A. E. L., Levis, C., Salles Machado, J., Bueno, L., and Peroni, N. (2020). Pre-colonial amerindian legacies in forest composition of Southern Brazil. *PLoS ONE* 15:e0235819. doi: 10.1371/journal.pone.0235819
- Pereira, G. S., Noelli, F. S., Bitencourt Campos, J., Pereira Santos, M., and Zocche, J. J. (2016). Ecologia Histórica Guarani: as plantas utilizadas no Bioma Mata Atlântica do litoral sul de Santa Catarina, Brasil (Parte 1). *Cadernos do LEPAARQ (UFPEL)* 13, 197–246. doi: 10.15210/lepaarq.v13i26.7608
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., and Blair, M. E. (2017). Opening the black box: an open-source release of maxent. *Ecography* 40, 887–93. doi: 10.1111/ecog.03049
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–97. doi: 10.1890/07-2153.1
- Piperno, D. R. (2011). The origins of plant cultivation and domestication in the new world tropics: patterns, process, and new developments. *Curr. Anthropol.* 52, S453–70. doi: 10.1086/659998
- Piperno, D. R., Ranere, A. J., Holst, I., and Hansell, P. (2000). Starch grains reveal early root crop horticulture in the panamanian tropical forest. *Nature* 407, 894–97. doi: 10.1038/35038055
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* 7, 217–40. doi: 10.5194/soil-7-217-2021
- POWO (2024). Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew.
- Prümers, H., Betancourt, C. J., Iriarte, J., Robinson, M., and Schaich, M. (2022). Lidar reveals pre-hispanic low-density Urbanism in the Bolivian Amazon. *Nature* 606, 325–28. doi: 10.1038/s41586-022-04780-4
- Pyles, M. V., Silva Magnago, L. F., Maia, V. A., Pinho, B. X., Pitta, G., de Gasper, A. L., et al. (2022). Human impacts as the main driver of tropical forest carbon. *Sci. Adv.* 8:eabl7968. doi: 10.1126/sciadv.abl7968
- Rafuse, D. J. (2021). A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina 7:6. doi: 10.5334/oq.97
- Reis, M. S., Ladio, A., and Peroni, N. (2014). Landscapes with *Araucaria* in South America: evidence for a cultural dimension. *Ecol. Soc.* 19. doi: 10.5751/ES-06163-190243

- Reis, M. S., Montagna, T., Mattos, A. G., Filippin, S., Ladio, A. H., Cunha Marques, A., et al. (2018). Domesticated landscapes in Araucaria Forests, Southern Brazil: a multispecies local conservation-by-use system. *Front. Ecol. Evol.* 6:00011. doi: 10.3389/fevo.2018.00011
- Riris, P. (2019). Sparse radiocarbon data confound culture-climate links in late Pre-Columbian Amazonia. *Quaternary* 2. doi: 10.3390/quat2040033
- Riris, P., and de Souza, J. G. (2021). Formal tests for resistance-resilience in archaeological time series. *Front. Ecol. Evol.* 9:740629. doi: 10.3389/fevo.2021.740629
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–29. doi: 10.1111/ecog.02881
- Roberts, P., Boivin, N., and Kaplan, J. O. (2018). Finding the anthropocene in tropical forests. *Anthropocene* 23, 5–16. doi: 10.1016/j.ancene.2018.07.002
- Roberts, P., Hunt, C., Arroyo-Kalin, M., Evans, D., and Boivin, N. (2017). The deep human prehistory of global tropical forests and its relevance for modern conservation. *Nat. Plants* 3, 1–9. doi: 10.1038/nplants.2017.93
- Roberts, P., Kaplan, J. O., Findley, D. M., Hamilton, R., Caetano-Andrade, V. L., Amano, N., et al. (2023). Mapping our reliance on the tropics can reveal the roots of the anthropocene. *Nat. Ecol. Evol.* 7, 632–36. doi: 10.1038/s41559-023-01998-x
- Robinson, M., De Souza, J. G., Maezumi, S. Y., Cárdenas, M., Pessenda, L., Prufer, K., et al. (2018). Uncoupling human and climate drivers of late holocene vegetation change in Southern Brazil. *Sci. Rep.* 8:7800. doi: 10.1038/s41598-018-24429-5
- Rojas Mora, S., and Gaitán, F. M. (2015). Análisis espacial del sitio arqueológico San Pedro, ubicado en el bajo río San Jorge, Caribe colombiano. *Revista Colombiana de Antropología* 51, 339–63. doi: 10.22380/2539472X24
- Scerri, E. M. L., Roberts, P., Maezumi, S. Y., and Malhi, Y. (2022). Tropical forests in the deep human past. *Phil. Transact. Royal Soc. B: Biol. Sci.* 377:20200500. doi: 10.1098/rstb.2020.0500
- Schmidt, M. J., Py-Daniel, A. R., de Paula Moraes, C., Valle, R. B. M., Caromano, C. F., Teixeira, W. G., et al. (2014). Dark earths and the human built landscape in amazonia: a widespread pattern of anthrosol formation. *J. Archaeol. Sci.* 42, 152–65. doi: 10.1016/j.jas.2013.11.002
- Schmitz, P. L., Rogge, J. H., Novasco, R. V., Mergen, M. N., and Ferrasso, S. (2013). Rincão dos Albinos - Um grande sítio Jê Meridional. *Pesquisas. Antropologia* 70, 65–131. doi: 10.5935/0553-8467.20170005
- Sillero, N., and Barbosa, A. M. (2021). Common mistakes in ecological niche models. *Int. J. Geogr. Inform. Sci.* 35, 213–26. doi: 10.1080/13658816.2020.1798968
- Silva, L. C. R., Wood, M. C., Johnson, B. R., Coughlan, M. R., Brinton, H., McGuire, K., et al. (2022). A generalizable framework for enhanced natural climate solutions. *Plant Soil* 479, 3–24. doi: 10.1007/s11104-022-05472-8
- South, A., Michael, S., and Massicotte, P. (2023). *Rnaturalearthhires: High Resolution World Vector Map Data from Natural Earth Used in Rnaturalearth*. Manual.
- Souza, A. F. (2021). A Review of the structure and dynamics of araucaria mixed forests in Southern Brazil and Northern Argentina. *New Zealand J. Botany* 59, 2–54. doi: 10.1080/0028825X.2020.1810712
- Steen, V. A., Tingley, M. W., Paton, P. W. C., and Elphick, C. S. (2021). Spatial thinning and class balancing: key choices lead to variation in the performance of species distribution models with citizen science data. *Met. Ecol. Evol.* 12, 216–26. doi: 10.1111/2041-210X.13525
- Toso, A., Hallingstad, E., McGrath, E., Fossile, T., Conlan, Jessica Ferreira, C., Bandeira, C. D., et al. (2021). Fishing intensification as response to late holocene socio-ecological instability in Southeastern South America. *Sci. Rep.* 11:23506. doi: 10.1038/s41598-021-02888-7
- Tulowiecki, S. J., and Larsen, C. P. S. (2015). Native American impact on past forest composition inferred from species distribution models, Chautauqua County, New York. *Ecol. Monogr.* 85, 557–81. doi: 10.1890/14-2259.1
- Tulowiecki, S. J., Ranney, E. R., Keenan, E. M., Neubert, G. M., and Hogan, M. L. (2022). Localized native american impacts on past forest composition across a regional extent in North-Eastern United States. *J. Biogeogr.* 49, 1099–109. doi: 10.1111/jbi.14369
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Aroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Met. Ecol. Evol.* 10, 225–32. doi: 10.1111/2041-210X.13107
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Aroita, G. (2021). Modelling species presence-only data with random forests. *Ecography* 44, 1731–42. doi: 10.1111/ecog.05615
- Valavi, R., Guillera-Aroita, G., Lahoz-Monfort, J. J., and Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* 92:e01486. doi: 10.1002/ecm.1486
- Wachtel, I., Zidon, R., Garti, S., and Shelach-Lavi, G. (2018). Predictive modeling for archaeological site locations: comparing logistic regression and maximal entropy in north israel and North-East China. *J. Archaeol. Sci.* 92, 28–36. doi: 10.1016/j.jas.2018.02.001
- Walker, R. S., Ferguson, J. R., Olmeda, A., Hamilton, A. J., Elghammer, J., and Buchanan, B. (2023). Predicting the geographic distribution of ancient amazonian archaeological sites with machine learning. *PeerJ* 11:e15137. doi: 10.7717/peerj.15137
- Wang, H., Caetano-Andrade, V., Boivin, B., Clement, C. R., Espindola Ayala, W., Dario Melinski, R., et al. (2025). Long-term human influence on the demography and genetic diversity of the hyperdominant *bertholletia excelsa* in the amazon basin. *Curr. Biol.* 35, 629–639.e4. doi: 10.1016/j.cub.2024.12.023
- Wesolowski, V., Mendonça de Souza, S. M. F., Reinhard, K. J., and Ceccantini, G. (2010). Evaluating Microfossil Content of Dental Calculus from Brazilian Sambaquis. *J. Archaeol. Sci.* 37, 1326–38. doi: 10.1016/j.jas.2009.12.037
- Whitford, A. M., Shipley, B. R., and McGuire, J. L. (2024). The influence of the number and distribution of background points in presence-background species distribution models. *Ecol. Model.* 488:110604. doi: 10.1016/j.ecolmodel.2023.110604
- Wieder, W. R., Boehnert, J., Bonan, G. B., and Langseth, M. (2014). RegridDED Harmonized World Soil Database V1.2. ORNL DAAC, September.
- Witteveen, N. H., Kleijwegt, Z. S., Geara, H., Kool, C., Blaus, A., Cabrera Saenz, L., et al. (2025). Quantifying past forest cover and biomass changes in the ecuadorian amazon. *New Phytol.* 245, 141–53. doi: 10.1111/nph.20237
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., et al. (2013). presence-only modelling using maxent: when can we trust the inferences? *Met. Ecol. Evol.* 4, 236–43. doi: 10.1111/2041-210X.12004
- Yu, B., Wenyu Dai, W., Li, S., Wu, Z., and Wang, J. (2024). A new threshold selection method for species distribution models with presence-only data: extracting the mutation point of the P/E curve by threshold regression. *Ecol. Evol.* 14:e11208. doi: 10.1002/ece3.11208