# Imaging spectroscopy predicts variable distance decay across contrasting Amazonian tree communities

Frederick C. Draper[1,2], Christopher Baraloto[2], Philip G. Brodrick[1], Oliver L. Phillips[3], Rodolfo Vasquez Martinez[4], Euridice N. Honorio Coronado[5], Timothy R. Baker[3], Ricardo Zárate Gómez[5], Carlos A. Amasifuen Guerra[6], Manuel Flores[6], Roosevelt Garcia Villacorta[7], Paul V. A. Fine[8], Luis Freitas[5], Abel Monteagudo-Mendoza [4,9], Roel J.W Brienen[3] and Gregory P. Asner[1]

[1] Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA

[2] International Center for Tropical Botany, Florida International University, Miami, FL, USA

[3] School of Geography, University of Leeds, Leeds, West Yorkshire, UK

[4] Jardín Botánico de Missouri, Oxapampa, Pasco, Perú

[5] Instituto de Investigaciones de la Amazonia Peruana, Iquitos, Loreto, Perú

[6] Facultad de Ciencias Forestales, Universidad Nacional de la Amazonía Peruana, Iquitos, Loreto, Perú

[7] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

[8] Department of Integrative Biology, University of California, Berkeley, CA, USA

[9] Universidad Nacional de San Antonio Abad del Cusco, Cusco, Perú

18

**Summary**

19 **Summary**

20 1. The forests of Amazonia are among the most biodiverse on Earth, yet accurately quantifying

21 how species composition varies through space (i.e. beta-diversity) remains a significant

22 challenge. Here we use high-fidelity airborne imaging spectroscopy from the Carnegie Airborne

23 Observatory to quantify a key component of beta-diversity, the distance decay in species

24 similarity through space, across three landscapes in Northern Peru. We then compared our

25 derived distance decay relationships to theoretical expectations obtained from a Poisson Cluster

26 Process, known to match well with empirical distance decay relationships at local scales.

27 2. We used an unsupervised machine learning approach to estimate spatial turnover in species

28 composition from the imaging spectroscopy data. We first validated this approach across two

29 landscapes using an independent dataset of forest composition in 49 forest census plots (0.1-1.5

30 ha). We then applied our approach to three landscapes, which together represented terra firme

31 clay forest, seasonally-flooded forest and white-sand forest. We finally used our approach to

32 quantify landscape-scale distance decay relationships and compared these with theoretical

33 distance decay relationships derived from a Poisson Cluster Process.

34 3. We found a significant correlation of similarity metrics between spectral data and forest plot

35 data, suggesting that beta-diversity within and among forest types can be accurately estimated

36 from airborne spectroscopic data using our unsupervised approach. We also found that estimated

37 distance decay in species similarity varied among forest types, with seasonally-flooded forests

38 showing stronger distance decay than white-sand and terra firme forests. Finally, we

39 demonstrated that distance decay relationships derived from the theoretical Poisson Cluster

40 Process compare poorly with our empirical relationships.

41 4. Synthesis: Our results demonstrate the efficacy of using high-fidelity imaging spectroscopy to

42 estimate beta-diversity and continuous distance decay in lowland tropical forests. Furthermore,

43 our findings suggest that distance decay relationships vary substantially among forest types,

44 which has important implications for conserving these valuable ecosystems. Finally, we

45 demonstrate that a theoretical Poisson Cluster Process poorly predicts distance decay in species

46 similarity as conspecific aggregation occurs across a range of nested scales within larger

47 landscapes.

48

49

**Introduction**

The forests of Amazonia are highly diverse, supporting as many as 16,000 tree species (ter

Steege et al., 2013). The importance of this diversity, beyond its intrinsic value as a natural

wonder, is increasingly well documented, for example, by underpinning key biogeochemical

cycles and determining the resilience of Amazonian forests to climate change (Sakschewski et al.

2016). Despite this recognition of the importance of diversity, accurately quantifying how

species composition varies through space (i.e., beta-diversity) in Amazonia remains a significant

challenge given the remoteness of the largest tropical forest on Earth. Over recent years large

networks of forest plots (e.g. RAINFOR, ATDN and CTFS) have provided invaluable insight

into the spatial ecology of Amazon forests (Duque et al., 2017; Phillips et al., 2004; ter Steege et

al., 2006). However, even summed together these networks represent only ~2000 ha of forest,

with many plots in localised clusters. Therefore, using plot data alone to assess continuous

spatial phenomena such as turnover in species composition represents a significant current

limitation to understanding tropical biodiversity.

An alternative, yet complementary, approach to quantifying biodiversity is through the use of

remotely sensed data integrated with existing plot data. Such an approach enables the acquisition

of contiguous data over vast swaths of forests irrespective of accessibility, potentially

transforming the power of an entirely ground-based approach. Multispectral data from satellite

based remote sensing, in conjunction with plot data, has been used successfully to broadly

classify different forest types (Draper et al., 2014; Salova et al., 2005) and to provide general

assessment of species turnover in Amazonia (Thessler, 2008; Tuomisto et al., 2003). However,

current satellite based multispectral sensors (e.g. Landsat) lack the spatial and spectral resolution

required to sufficiently differentiate the high species-level diversity occurring within tropical

forests (Rocchini, 2007a, 2007b; Rocchini et al., 2016). Recent advances in high-fidelity, laser

guided imaging spectroscopy present a viable solution, and have been used successfully to

estimate beta-diversity in Neotropical forests (Féret & Asner 2014a; b; Somers et al. 2015).

A key component of beta-diversity is the variation in species composition as a function of

geographic distance (hereafter referred to as distance decay). Distance decay is a particularly

useful concept as it allows for an understanding of the relative importance of different processes

3

79   that may determine patterns of beta-diversity, such as environmental filtering and dispersal

80   limitation (Soininen, McDonald, & Hillebrand 2007; Tuomisto, Ruokolainen, & Yli-Halla 2003).

81   Understanding variation in distance decay relationships among different landscapes and forest

82   types also has important implications for designing effective conservation strategies (Socolar et

83   al. 2016). For example, the gradient of distance decay can help to understand if conserving

84   species in a given landscape or forest type will be maximised by many small or few large

85   protected areas (Nekola & White 1999). Furthermore, distance decay relationships can be used to

86   formally test theoretical predictions of community assembly, for example from neutral theory

87   and sampling area theory (Chave & Leigh, 2002; Condit et al., 2002; Hubbell, 2001; Morlon et

88   al., 2008).

89   One particularly significant theoretical model suggests that distance decay relationships are

90   defined by the spatial aggregation of tree species, which can be characterized by a Poisson

91   Cluster Process (hereafter PCP) (Plotkin et al. 2000; Morlon et al. 2008). This model is useful

92   because it correctly recognises that tree species are spatially aggregated (Condit et al. 2000), but

93   does not attempt to ascribe a particular community assembly mechanism. Furthermore, this

94   model has accurately characterised species area curves, and distance decay relationships in

95   number of tropical forests (Plotkin et al. 2000; Morlon et al. 2008). Importantly, while this

96   model has found relatively good agreement at small scales ($\leq$ 50 ha), it has not been possible to

97   test this model at larger spatial scales. An key limitation of the PCP approach is that it assumes a

98   single scale of aggregation, in this paper we test the validity of this assumption at larger spatial

99   scales (>1000 ha).

100  Within western Amazonia, several plot based studies have examined distance decay relationships

101  in tree communities, and most of these studies find an initial rapid decay in species similarity

102  over the first few kilometres followed by a far more gradual decay over greater distances (Condit

103  et al., 2002; Duque et al., 2009; Tuomisto et al., 2003). However, this relationship varies

104  substantially with the spatial scale of study (Morlon et al., 2008; Phillips et al., 2003; Tuomisto

105  et al., 2003), forest type (Draper et al., 2018), underlying geology (Phillips et al., 2003) and

106  taxonomic group (Kristiansen et al., 2012; Tuomisto et al., 2003). Importantly, all of these plot

107  based studies have been data limited, either using a relatively small number of plots (typically <

108  50 ha) to interpolate distance decay over tens to hundreds of kilometres (Condit et al., 2002;

4

109 Tuomisto et al., 2003), or using spatially continuous data to investigate distance decay over small

110 spatial scales ($\leq$ 50 ha) (May et al., 2016; Morlon et al., 2008).

111 Here we apply a sequence of unsupervised machine learning techniques (Féret & Asner 2014b)

112 to continuous high-fidelity spectral datasets to quantify contiguous beta-diversity and associated

113 distance decay relationships at a landscape scale (>1000 ha) across three lowland landscapes in

114 Amazonian Peru. At each of these landscapes we apply our method to one of three distinct forest

115 types: white-sand forest, seasonally-flooded forest and terra firme clay forest. We also use an

116 extensive network of 49 forest census plots across two landscapes to thoroughly validate our

117 approach and to answer the following questions.

118     1. Does high spatial resolution imaging spectroscopy accurately predict turnover in tree
119        species composition across different forest types in lowland Amazonia?

120     2. How does distance decay in tree species composition vary across different forest types in
121        lowland Amazonia?

122     3. How well does a theoretical PCP predict distance decay in tree species composition across
123        a range of forest types in lowland Amazonia?

124

125

**Methods**

Study landscapes

Three distinct landscapes were used in this study: Allpahuayo Mishana, Jenaro Herrera, and

Quebrada Braga. These landscapes are all located in the department of Loreto, Peru (Fig. 1), and

were selected because they harbour at least one of the three most common forest types

encountered across western Amazonia: terra firme clay forest, seasonally-flooded forest, and

white-sand forests (Baraloto et al. 2011). The first landscape, Allpahuayo Mishana, is a national

reserve located close to the city of Iquitos that contain a mosaic of terra firme clay and white-

sand forest (Fine et al., 2010; García Villacorta et al., 2003). These white-sand forests have

exceptionally nutrient poor sandy soils of cratonic origin, and harbour numerous endemic tree

species (Fine et al. 2010). The second landscape, Jenaro Herrera, is a centre of research of the

Instituto de Investigaciones de la Amazonía Peruana (IIAP). Jenaro Herrera is made up primarily

of terra firme forest, although there are some small patches of white-sand forest, seasonally-

flooded forest, and palm swamp forest (Honorio Coronado et al., 2008; Honorio Coronado et al.,

2009). Finally, the Quebrada Braga landscape is located south of Jenaro Herrera, and is

surrounded by the Ucayali river on three sides, these low-lying forests are inundated seasonally

with nutrient-rich white water (Nebel et al. 2001).

Airborne data

We used the Carnegie Airborne Observatory (CAO) Airborne Taxonomic Mapping System

(AToMS) to obtain fused high fidelity imaging spectroscopy and Light Detection and Ranging

(LiDAR) data for all three of our landscapes (Asner et al. 2012). CAO flights took place between

June and September 2012 at an altitude of approximately 2000 above ground level, with an

average flight speed of 60 m s$^{-1}$, and a mapping swath of ~1.2 km. Spectral radiance data were

collected between 380 and 2510 nm at 5 nm increments (Asner et al. 2012). These measurements

were subsequently resampled to 10-nm resolution, resulting in 214 contiguous spectral bands at a

ground-level resolution (pixel size) of 2 m. LiDAR data were obtained from a dual laser

waveform scanner that was operated at 200 kHz, with a 17° scan half-angle from nadir, yielding

a point density of 4 laser shots m$^{-2}$ (up to 16 returns m$^{-2}$). Lidar data were used to produce maps

of tree canopy height and ground surface at 1-m spatial resolution. Spectral and LiDAR data

6

155 were precisely geo-located using an embedded high resolution Global Positioning System-
156 Inertial Measurement Unit (GPS-IMU).

157 The spectral radiance data were atmospherically corrected to apparent surface reflectance with
158 the ACORN-5 model (Imspec LLC, Glendale, CA USA). Images were then processed to exclude
159 pixels that were not fully sunlit (i.e. shaded by another tree), covered by cloud, or represented a
160 non-forested land surface. Shade masks were built using LiDAR-derived ray tracing models
161 (Asner et al. 2007), clouds were masked manually, and non-forested land surfaces were
162 identified using a LiDAR derived map of tree canopy height where pixels with a canopy < 3 m
163 were considered non-forested. In addition, spectral bands that contained sampling noise
164 (wavelengths < 400 nm and > 2500 nm) or that were dominated by atmospheric water vapour
165 (wavelengths 1350-1480 nm and 1780-2032 nm), were not used in this analysis.

166 Estimating beta-diversity from spectral data

167 To estimate beta-diversity from spectral data, we used the 'spectral species distribution' (SSD)
168 approach, building on the previous work of Féret & Asner (2014 a, b) and more generally on the
169 foundations of the spectral variation hypothesis (Palmer et al., 2002). Our approach assumes that
170 the spectral properties of a landscape vary with species composition, and therefore we are able to
171 use variation in spectral composition as a proxy for variation in species composition. At each of
172 the three sites, we independently applied a seven-step analysis procedure to generate our mapped
173 estimates of tree species compositional change as follows.

(1) We performed a principal component analysis (PCA) on our processed spectral image in
175 order to reduce the high dimensionality of the spectral data and to isolate and remove
176 sampling artefacts such as cross-track brightness gradients.

(2) We manually selected components associated with biological gradients by visually
178 examining the first 35 components, and removing any that showed obvious artefacts,
179 such as clear striping. This left 4-8 useful components that were used in steps 3-7. At all
180 landscapes the first three components were always selected and the together the
181 components represented >60% of the variance.

(3) We applied k-means clustering to the selected components, clustering each pixel into one
183 of 50 possible 'spectral species'. Spectral species being simply clusters of pixels that
184 have similar reflectance values, which may, but equally may not, trace onto actual

185       species. This process reduces the multi-layer image of PCs into a single layer image

186       containing the spatial distribution of spectral species. Due to the large size of the dataset,

187       k-means was applied using the 'mini-batch k-means' function in the Python package

188       scikit learn, which provides near-equivalent performance at rapid computational speed

189       (Pedregosa et al. 2012). Mini-batches of 10,000 pixels were used, each with 20 random

190       starts.

191   (4) We then divided the resulting spectral species distribution image into 1 ha mapping

192       kernels. Kernels in which > 66 % of pixels corresponded to either shade, non-vegetated

193       ground, or were clouded were excluded from all further analysis. This led to a ~20 % loss

194       of area from each landscape (Table 1).

195   (5) We then converted the image into a spectral species abundance matrix where each row

196       corresponded to an individual kernel and each column to a spectral species, from which

197       we calculated a Bray-Curtis distance matrix.

198   (6) We then applied Non-metric multi-dimensional scaling (NMDS) to the distance matrix in

199       order to extract the most important compositional gradients in the spectral species data.

200       The NMDS was optimized for three axes and run for 30 iterations.

201   (7) Finally, we re-projected the three NMDS axis scores into a raster format so that spatial

202       variation in spectral species composition could be visualized.

203   The PCA and k-means analysis were undertaken using the Python package Sci-kit learn

204   (Pedregosa et al. 2012). All beta-diversity analyses (steps 5 and 6) were performed in the R

205   statistical environment using the Vegan package (Oksanen et al. 2013).

206   Plot inventory beta-diversity estimates

207   To validate our approach, we compared our estimates of beta-diversity derived from spectral data

208   to measured beta-diversity obtained from inventory plot data at Allpahuayo Mishana and Jenaro

209   Herrera. Our plot dataset consisted of 37 existing forest inventory plots distributed across

210   Allpahuayo Mishana in white-sand and terra firme forest types, and 12 forest plots distributed

211   across Jenaro Herrera in terra firme, white-sand and palm swamp forest types (Fig. 2). Plots

212   varied in size from 0.1-1.5 ha, and five different sampling protocols were used as described

213   below.

214    We used 12 large rectangular permanent sampling plots (0.5 to 1.5 ha), in which all tree stems

215    with a diameter greater than 10 cm have been tagged and identified. Seven of these rectangular

216    plots were one ha in size and belong to the RAINFOR Network, two of these plots were 1.5 ha in

217    size (Peacock et al., 2007; Martinez & Phillips 2000). We also used three rectangular 0.5 ha plots

218    in which all stems greater than 5 cm have been identified (Honorio Coronado et al., 2008). We

219    further used 16 small 0.1 ha plots, in which all stems greater than 2.5 cm in diameter were

220    identified. Six of these 0.1 ha plots were 'Gentry' plots consisting of ten 2 x 50 m intersecting

221    transects (Gentry, 1982; Phillips et al., 2003). These six Gentry plots, alongside the seven 1 ha

222    RAINFOR plots were downloaded from the ForestPlots.net online repository (Lopez-Gonzalez

223    et al., 2009; Lopez-Gonzalez et al., 2011). The ten remaining 0.1 ha plots were rectangular 20 x

224    50 m plots (Zárate et al., 2006). We used four 0.5 ha modified Gentry plots, within which all

225    stems greater than 2.5 cm in diameter were identified (Baraloto et al. 2011). The remaining 14

226    plots were circular plots in which all species greater than 10 cm dbh were identified (Baldeck et

227    al., 2016); two of these circular plots were 0.25 ha and 12 were 0.14 ha. Summary details of the

228    inventory plot dataset are given in Table 2, and full details of all plots are given in table S.1.

229    GPS coordinates were taken in the centre of each plot to determine its position within the

230    landscape. There are significant uncertainties associated with using a GPS underneath a forest

231    canopy, particularly for smaller inventory plots. Our approach partially mitigates these

232    uncertainties as our aim is to align these plots with spectral species composition estimates at a 1

233    ha scale, and therefore, GPS locations need only be located in the correct 1 ha kernel. Ultimately,

234    we removed five plots from this aggregate dataset in Allpahuayo Mishana (four 0.1 ha and one

235    0.5 ha), that were located < 10 m from a kernel boundary between white-sand forest and terra

236    firme forest according to our spectrally derived map of estimated beta-diversity. As these plots

237    were larger than 10 m in any dimension, there is a high likelihood that much of the area of these

238    plots was situated in an incorrect kernel. These five boundary plots introduced additional

239    variation in the relationship, as shown in Figure 3.

240    Because morpho-species were not standardised across datasets, it was necessary to exclude all

241    individuals not identified to species level from the dataset before calculating beta-diversity.

242    These exclusions led to a loss of 5-20% of individuals, which is likely to slightly increase the

243    similarity among plots. However, patterns of beta-diversity among Amazonian tree census plots

244 have been shown to be generally robust to the exclusion of similar proportions of morpho-
245 species (Pos et al. 2014).

246 Given that estimates of beta-diversity are sensitive to the number of individuals per plot, and that
247 our dataset was made up of plots of different sizes (and different numbers of individuals), it was
248 necessary to standardise our plot dataset by stem number before calculating beta-diversity. We
249 did this by using a bootstrap resampling process. This process consisted of first establishing the
250 minimum number of individuals in any plot, in this case 65, and then sampling (without
251 replacement) 65 individuals from each plot at random. A Bray-Curtis distance matrix was then
252 constructed using this subsample of 65 individuals per plot. Using this distance matrix, NMDS
253 ordinations were performed. NMDS axis scores were then extracted for each plot. This process
254 was then repeated 1000 times with a different set of 65 individuals per plot in order to develop
255 confidence intervals for NMDS axis scores. Finally, we were able to compare NMDS axis scores
256 derived from this plot inventory data with the corresponding NMDS axis scores derived from the
257 spectral data.

258 Estimating spectral distance decay

259 To estimate the distance decay in species composition from spectral data within forest types, it
260 was first necessary to isolate pixels that correspond to the forest type of interest. At Allpahuayo
261 Mishana, the target forest type was white-sand forest. Using our validation data, we
262 demonstrated that at this site white-sand forests can be readily separated from terra firme forests
263 based on spectral composition (Fig. 2 and 3). Therefore, pixels with a value of greater than 0.3
264 on the first NMDS axis were classified as white-sand forest.

265 At Jenaro Herrera the target forest type was terra firme forest. We first used our spectral data to
266 exclude small patches of white-sand forest from our analysis; to do this, we excluded all pixels
267 with a value of greater than 0.2 on the second NMDS axis as this was shown to represent white
268 sand forests in the validation data (Fig. 3). We then used the LiDAR derived DEM to separate
269 pixels of seasonally-flooded forest from terra firme forest. Kernels with a mean elevation greater
270 than 118 m were considered to be terra firme forest.

271 At Quebrada Braga the target forest type was seasonally-flooded forest. We used our LiDAR
272 derived DEM to isolate those forests that are seasonally-flooded from those that are not. We

10

273    were able to use existing plot data to identify the elevation of seasonally-flooded forests (Kvist &

274    Nebel 2001; Nebel et al. 2001). All kernels that had a mean elevation of 113-117 m a.s.l. were

275    deemed to be seasonally-flooded. As this landscape is surrounded on three sides by a white-

276    water river, we assume that seasonal flooding provides uniformly high nutrient deposition and

277    that there are no further edaphic gradients.

278    To visualise the distance decay across each landscape we calculated the mean similarity (inverse

279    Bray Curtis) for all paired plots within bins of 100 m, (i.e. the mean similarity between plots

280    located 0-100 m apart, 100-200 m apart etc.). We have presented the ensemble mean and

281    standard deviation with each distance bin and do not assume independence among these pairwise

282    distances. Additionally, we calculated the first order derivative of similarity every 100 m across

283    each landscape. We used a LOESS smoothing function (span = 0.35), to demonstrate how the

284    derivative varies with distance across each landscape.

285    Theoretical distance decay

286    To assess the extent to which our empirical spectral distance decay relationships could be

287    reproduced by a PCP, we applied the theoretical framework outlined by Morlon et al. (2008).

288    Because we applied this approach to 50 spectral species rather than hundreds or thousands of

289    species, it was essential that our measure of similarity was calculated using abundance rather

290    than occurrence data. Therefore, we did not fit the general formula supplied by Morlon et al.

291    (2008) which had been developed to using the Sorensen index. Instead, we simulated maps of

292    spectral species distributions with a PCP, which we parameterised using fits of Ripley's K curves

293    to our spectral species maps. Subsequently, we were able to derive abundance-based distance

294    decay relationships from these theoretically derived maps of spectral species distributions.

295    The PCP is a stochastic mathematical process of assigning clusters of objects (here spectral

296    species) in space according to the following: 1. Cluster centres for each object are randomly

297    distributed across a landscape assuming a constant cluster density. The number of individuals in

298    each cluster is drawn from a Poisson distribution. 3. Individuals within each cluster are then

299    distributed based on a radially symmetrical Gaussian distribution .

300    In this study, a PCP was produced for each of the 50 spectral species across each of the three

301    landscapes according to the following process:

302      (1) Empirical Ripley's K curves were derived for each spectral species in each landscape

303          using the R package Spatstat (Baddeley & Turner 2005).

304      (2) When a Ripley's K curve is calculated for a PCP, it can be shown to have the functional

305          from presented in Equation 1 (Plotkin et al. 2000). Consequently, we use an inverse

306          modelling framework to match each empirically derived Ripley's K curve with Equation

307          1 by adjusting $\rho$ (the density of clusters across the landscape), and $\mu$ (the intensity of

308          individuals within each cluster).

309          Equation 1: $K(d)^{PCP} = \pi d^2 + \rho^{-1}\left(1 - \exp\left(\frac{-d^2}{4\mu^2}\right)\right)$

310      (3) Species likelihood probabilities were then determined for each spectral species using the

311          $\rho$ and $\mu$ values in a PCP in concert with the radial Gaussian probability function defined

312          in Equation 2. Probabilities from each clump were overlaid on top of one another and the

313          maximum likelihood was used.

314          Equation 2: $h(x, y) = (2\pi\mu^2)^{-1} \exp\left(\frac{-(x^2+y^2)}{2\mu^2}\right)$

315      (4) The 50 species likelihood maps (one per spectral species) were then normalized based on

316          the abundance of each spectral species in the empirical maps. These likelihoods were

317          then used to weight a random draw that was used to condense the likelihoods into a

318          single, theoretically-based spectral species map.

319      (5) A one ha grid was then fit over the simulated spectral species distribution map and the

320          Bray Curtis distance among one ha kernels was calculated in exactly the same way as

321          was done with the empirical data. From this grid, theoretical distance decay relationships

322          were calculated in exactly the same manner as was done with the empirical spectral data

323          (i.e. by calculating the mean similarity (inverse Bray Curtis) for all paired plots within

324          bins of 100 m.

325      (6) Steps 2 to 5 were then repeated 20 times, to generate 20 distinct theoretical spectral

326          species maps and associated distance decay curves. The final curves presented were the

327          mean of means within each 100 m bin and the standard deviations of the means.

328

## Results

Validation with forest plot data

At Allpahuayo Mishana, our estimates of species compositional turnover derived from spectral data were strongly correlated with field plot-based measures of beta-diversity ($R^2 = 0.85$; $P < 0.001$; Fig. 3). However, the residual variance was higher among only terra firme forest plots ($R^2 = 0.29$; $P = 0.05$) than among only white-sand forest plots ($R^2 = 0.76$; $P < 0.001$). At Jenaro Herrera, there was also a highly significant relationship between beta-diversity estimated with our spectral approach and field-measured beta-diversity ($P = <0.001$), although there was more residual variance at this site than at Allpahuayo Mishana ($R^2=0.68$). Most of the variation in the relationship between spectral and plot data came from palm swamp forests, which were poorly distinguished in the second NMDS axis; instead, the third NMDS axis was more useful at identifying areas of palm swamp (figure S.2). The relationship between spectral composition and species composition was consistent across two landscapes, and among different field plot datasets that were established using different sampling protocols with different stem diameter size limits.

Mapping beta-diversity

Our spectrally-derived maps of estimated tree species composition demonstrate clear gradients across the three study landscapes (Fig. 4). However, the underlying determinants of these floristic gradients appear to be different among the three sites. At Allpahuayo Mishana, the three NMDS axes show similar spatial patterns (Figs 4 and S.1), with NMDS axes 2 and 3 additionally containing a substantial element of sampling artefact (i.e. clear striping). This relative uniformity across NMDS axes suggests there is a single predominant floristic gradient at this site, because, if multiple important floristic gradients were present, we would expect them to be reflected in different NMDS axes. Combined with field validation data, our spectrally-derived maps indicate that the primary floristic gradient at this site reflects an underlying edaphic gradient from nutrient rich terra firme clay soils, to nutrient poor white-sand soils. These white-sand forests were always found at higher elevations (>145 m a.s.l.) at Allpahuayo Mishana.

356  Our estimates of tree species composition also suggest that there is a strong spatial gradient in

357  floristic composition at Quebrada Braga. Similar to Allpahuayo Mishana, consistency among

358  NMDS axes suggests there is a single primary floristic gradient at Quebrada Braga (Fig. 4 and

359  S.2). Somewhat surprisingly, this floristic gradient did not correspond strongly with elevation.

360  The Quebrada Braga landscape is seasonally flooded by the large and nutrient-rich Ucayali

361  River, which surrounds this landscape on three sides. Therefore, elevation will primarily

362  determine the intensity and duration of this seasonal flooding.

363  Jenaro Herrera appears to be a more complex landscape than the other two, as it contains three

364  distinct floristic gradients, demonstrated by three distinctive NMDS axes (Fig. 4 and S.3). This

365  landscape appears to contain two forms of flooded forest, one flooded by nutrient-rich white

366  water from the large Ucayali River and another flooded by nutrient-poor black water. In addition,

367  there are patches of white-sand forest as well as forests that have been significantly impacted by

368  anthropogenic activities.

369  Empirical spectral distance decay

370  We observed a consistent pattern of a rapid decline in floristic similarity over distances of 500 m

371  or less across all three forest types. Beyond this initial steep decay in similarity, three patterns

372  distinguish these landscapes. In white-sand forests at Allpahuayo Mishana, after a rapid decay in

373  similarity over the initial 800 m there was almost no discernible decrease in similarity with

374  increasing distance (Fig. 5 panels A and D).

375  In seasonally-flooded forests at Quebrada Braga, we found a constant decay in floristic similarity

376  with increasing distance. As with the other two landscapes, this decline was steepest over the

377  initial 700 metres. However, the decline in compositional similarity persisted over the entirety of

378  this landscape, as demonstrated by the consistently negative differential values (Fig. 5 panels B

379  and D).

380  Finally, in terra firme forests at Jenaro Herrera we found a steep decay in compositional

381  similarity over 500 m, followed by a more gradual decline up to distances of 3 km (Fig. 5 panels

382  C and D). Beyond 3 km there was no discernible decrease in similarity with increasing distance

383  up to 10 km. Additionally, at Jenaro Herrera there was greater overall variation in compositional

384  similarity across all distances compared with the other two sites, as shown by the wider error

14

385    bars. We attribute this variation to the greater environmental variation at this site, as well as

386    greater overall species diversity in terra firme forests as opposed to both white-sand forests and

387    seasonally-flooded forests.

388    Theoretical distance decay model

389    Overall the theoretical models derived from our PCP approach poorly represented the three

390    empirical (spectrally derived) distance decay relationships (Fig. 5). At Allpahuayo Mishana,

391    although the form of the theoretical distance decay relationship was very similar to that derived

392    from the empirical data, the theoretically derived distance decay generally overestimates

393    similarity relative to the empirical data (Fig. 5 panel A). Similarly, Fig. 5 panel B shows that at

394    Jenaro Herrera, the general pattern of the distance decay relationship was reasonably

395    characterized relative to the empirical relationship, but the overall distance magnitude was not.

396    At Quebrada Braga, we found a very different pattern, with the PCP models predicting a

397    sustained sharp decrease in similarity over the first kilometre, which was not reflected in the

398    empirical data (Fig. 5 panel C). However, the shallow but continuous decline in similarity

399    beyond the first kilometre demonstrated by the PCP at Quebrada Braga showed reasonable

400    agreement with the empirically-based relationship (Fig. 5 panel C).

401

402

403 **Discussion**

404 Our results demonstrate that distance decay relationships vary among forest types in lowland

405 Amazonia at a landscape scale. This is significant, because in contrast with previous plot-based

406 studies, we are able to investigate this distance decay relationship continuously across landscapes

407 while simultaneously maintaining high resolution. Within terra firme forests, our estimated

408 distance decay curves are broadly consistent with a number of previous studies in this region

409 (Condit et al., 2002; Duque et al., 2009), showing both rapid decay in similarity over short

410 distances, followed by almost no decay at distances greater than 4 km. The two other forest types

411 that we investigated also demonstrate this initial rapid decline in similarity over the first

412 kilometre, supporting the idea that canopy tree species across forest types are spatially

413 aggregated over scales less than one kilometre (Condit et al. 2000). However, beyond this first

414 kilometre, patterns of distance decay sharply differ among different forest types.

415 The variation in distance decay among forest types is particularly apparent in seasonally-flooded

416 forest, which shows a strong and relatively continuous decline in similarity with increasing

417 distance. There are few plot based estimates of distance decay relationships in seasonally-

418 flooded forests with which to compare our data (but see Wittmann et al., 2006; Draper et al.,

419 2018). Nevertheless, our broad pattern of continuous decline in similarity appears to be

420 consistent with these plot-based analyses. Much of the variation in spectral species composition

421 across the Quebrada Braga landscape appears to be broadly independent of elevation. As

422 elevation here should be a reasonable proxy for flooding duration and intensity, our data suggest

423 that flooding duration and intensity are not the most important determinant of species

424 composition in this landscape. This contrasts with a number of previous studies that have found

425 flooding depth and duration to be the most important determinants of species composition (Assis

426 et al., 2015; Junk et al., 2011; Wittmann et al., 2004; Wittmann et al., 2006).

427 Instead, our results appear to emphasize the importance of disturbance in determining species

428 composition at this site. Disturbance has been recognised as an important driver of beta-diversity

429 in West Amazonian floodplain forests (Puhakka et al., 1992; Salo et al., 1986). This may be

430 especially true in Quebrada Braga as it is surrounded by the large and dynamic Ucayali River,

431 which migrates laterally over decadal timescales (Salo et al., 1986; Schwenk et al., 2017).

16

432    Therefore, while some areas might have experienced large-scale disturbance relatively recently,

433    other areas may not have been disturbed for many decades or centuries. Such disturbance

434    patterns would also be spatially auto-correlated, and therefore consistent with the distance decay

435    patterns we observe. The discrepancy between our study and previous plot-based studies (e.g.

436    Assis et al., 2015; Junk et al., 2011; Wittmann et al., 2004; Wittmann et al., 2006) may arise

437    from plot-based studies sampling predominantly mature seasonally-flooded forests over

438    disturbed forests, whilst our study samples the whole landscape without this apparent bias.

439    Fluvial disturbance is not the only form of large-scale spatially auto-correlated disturbance that

440    may be driving beta-diversity patterns in Amazonian forests; for example, in central Amazonia

441    large blow-down events have an important role in driving turnover in species composition

442    (Marra et al. 2014).

443    In white-sand forests, the initial rapid decline in similarity with increasing distance is even more

444    pronounced than in the other forest types and does not persist beyond the initial 800 m. This

445    initial rapid decay may reflect the patchiness of white-sand forests at Allpahuayo Mishana.

446    Patches of white-sand forests at this site are frequently smaller than 800 metres across, and

447    ecological similarity is likely to be higher within a patch than between patches. In this way,

448    white-sand forest tree communities may be functioning as meta-communities, separated by terra

449    firme forests (Adeney et al., 2016; Palacios et al., 2016). The lack of declining similarity with

450    increasing distance beyond 800 m is consistent with some published distance decay curves for

451    white-sand forests in this region (Draper et al., 2018), whist others that have been developed for

452    much broader spatial scales appear to show a more constant decay (García-Villacorta et al.,

453    2016; Guevara et al. 2016), presumably because they include several compositionally distinct

454    floras.

455    Jenaro Herrera presents a different, and perhaps more complex pattern than in the other

456    landscapes, indicated by the three NMDS axes showing distinct spatial patterns that reflect

457    different underlying gradients. For example, patches of white-sand forests and terra firme forests

458    are clearly distinct in NMDS axis 2, whilst palm swamp forests appear more strongly in the third

459    NMDS axis. Furthermore, unlike the other two landscapes, Jenaro Herrera appears to show a

460    strong anthropogenic disturbance gradient, which can be seen in high values in NMDS axis 1

461    that cluster near the town (fig. S.1). This apparently high level of anthropogenic disturbance is in

462    some ways unsurprising as Jenaro Herrera supports a larger population than the other two sites

463    and is surrounded by forests that are accessible and without formal legal protection. This

464    contrasts with the other two landscapes, with Allpahuayo Mishana being accessible but protected

465    and Quebrada Braga being unprotected but further from human development and due to seasonal

466    flooding, relatively inaccessible.

467    A clear feature revealed by our LiDAR-derived DEM at Jenaro Herrera is the sharp increase in

468    elevation that bisects the landscape from West to East (Fig. 4). This geological feature appears to

469    be a boundary between the upland Tertiary Iquitos geanticline and Pleistocene alluvial terraces

470    (Dumont et al., 1990; Dumont et al., 1991; Rasanen et al., 1992). Interestingly, this boundary

471    appears to have little impact on floristic composition unlike other geological features in this

472    region (Higgins et al., 2011, 2012). While field data will be required to confirm that there is little

473    floristic turnover across this boundary, the boundary does not appear in local floristic

474    classifications nor in maps of forest types (López Parodi & Freitas 1990; Honorio et al. 2008).

475    We were able to validate our approach by comparing our spectrally-derived estimates of beta-

476    diversity with an extensive network of 53 forest plots distributed across two sites. Overall, this

477    comparison provides compelling evidence that high fidelity imaging spectroscopy can be used to

478    understand the spatial organisation of biodiversity in hyper-diverse tropical forests. Our results

479    show highly significant linear relationship between spectrally-derived and plot-based estimates

480    of beta-diversity consistent with previous studies that have used similar unsupervised approaches

481    (Baldeck & Asner, 2013; Féret & Asner, 2014a, 2014b; Somers et al., 2015). Importantly, this

482    strong relationship is preserved across plots using both 2 cm and 10 cm diameter cut-offs. As the

483    spectral signal is derived entirely from the uppermost canopy layer, our results suggest that

484    canopy level species composition may an excellent proxy for species composition in understory

485    strata in these landscapes. The weaker relationship between spectral similarity and floristic

486    similarity in terra firme forests may reflect the fact that fewer canopy species were recorded in

487    this forest type. This is because the majority of stems recorded in the 0.1 ha plots are < 10 cm

488    dbh, which will not reach the forest canopy in these tall forests. In the shorter stature white-sand

489    forests, a larger proportion of small-stemmed trees will reach the canopy and therefore will be

490    included in the spectral data.

18

491    Across all forest types, the distance decay relationships derived from the theoretical PCP

492    compared poorly with the comparable empirical data. This mismatch suggests that the decay in

493    community composition cannot be easily predicted by the clustering of conspecific individuals

494    following a PCP. Major limitations of the PCP approach include the assumption that conspecific

495    individuals are aggregated at a single scale, and the assumption that each clump of individuals

496    throughout the landscape has the same Gaussian dispersal pattern (Morlon et al. 2008). The

497    single scale of aggregation assumption may be largely correct at small spatial scales ($\leq$ 50 ha) in

498    relatively homogenous environments (Morlon et al. 2008), where trees are aggregated mainly at

499    small scales < 50 m (Condit et al. 2000). However, at larger spatial scales (> 500 ha), conspecific

500    individuals aggregate at a range of different scales due to dispersal limitation, environmental

501    specificity, Janzen-Connell effects, and competition among individuals (Levin, 1992; Wiegand et

502    al., 2007). Similarly, a species is unlikely to have constant density across a 50 ha plot, however

503    across a landscape > 1000 ha, assuming a constant density becomes an even less plausible

504    assumption. Our results demonstrate that these assumptions would need to be relaxed in order to

505    reasonably predict distance decay relationships a landscape scales from theoretical spatial point

506    process models such as the PCP.

507    Furthermore, our theoretical approach calculates PCP distributions for each spectral species

508    independently; these distributions are then combined into a single map using random draws

509    weighted by the landscape abundance of each spectral species in the empirical spectral species

510    map. Our approach does not include interactions among species and between species and the

511    environment, instead assuming the landscape is a homogeneous plane. Incorporating these biotic

512    and abiotic interactions in future models could provide a way to further explore the relative

513    influence of neutral and niche processes at landscape scales. Finally, our PCP was parameterized

514    with by spectral species distributions. It is possible that parameterization based on actual species

515    distribution data, which would be extremely difficult to collect at such large scales, may lead to

516    different results. An approach integrating field and spectral species distributions could provide

517    further insight.

518    A more general limitation of our approach is that we cluster the spectral signal of the entire

519    landscape into just 50 spectral species and assume they are representative of hyper-diverse

520    tropical forest landscapes that will contain hundreds (if not thousands) of tree species. While this

521  approach is well supported by both our comparisons with field data, and previous work that has

522  shown 40 spectral species to be optimal (Feret & Asner 2014a), there are limitations. In general,

523  it is likely that common canopy species will dominate the spectral signal as they make up a far

524  greater proportion of the sunlit canopy, whilst rare and or understory species will be under-

525  represented. Rare species are thought to have more localised and environmentally specific

526  distributions (Hubbell, 2013), and therefore, the extent to which common species can be used to

527  investigate spatial patterns of beta-diversity merits further investigation.

528  Additionally, using 50 spectral species elevates the similarity among plots within each forest

529  type. This is especially evident in white-sand forests, where overall similarity is far higher in our

530  spectral based analysis than has been found previously in plot based studies (Fine et al., 2010;

531  García-Villacorta et al., 2016; Guevara et al., 2016, Draper et al., 2018). Many white-sand

532  specialist tree species share functional characteristics that are likely to make them spectrally

533  similar, such as increased leaf thickness and toughness, as well as lower concentrations of foliar

534  N and P (Asner et al., 2016; Fortunel et al., 2014; Fyllas et al., 2009). Therefore, the diversity

535  within white-sand forests may be poorly represented by our approach, resulting in an artificial

536  increase in similarity between plots. However, the tight correlation between spectral and plot-

537  based estimates of species composition in white-sand forests suggests that despite the overall

538  increase in similarity among plots, our approach is still able to capture the main correlates of plot

539  diversity.

540  The strength of our approach is that we can apply this method continuously to much larger areas

541  than would be impossible using field data alone. Therefore, there is great potential for using our

542  method to quantify beta-diversity and distance decay relationships continuously over far greater

543  spatial extents. Furthermore, our approach is not only able to quantify beta-diversity, but also to

544  precisely geo-locate where turnover occurs and therefore to suggest which environmental

545  features may be important. We suggest that unsupervised spectral-based approaches, such as

546  ours, can be used to actively guide field efforts to areas containing floristic assemblages that are

547  poorly represented by current plot networks. We advocate for closer collaboration among

548  ecologists using field-based data and those using imaging spectroscopy data.

549  In summary, this study demonstrates that distance decay relationships vary substantially among

550  landscapes and forest types in lowland Amazonia, consistent with much of what has been found

551     previously using field plot-based data. Nevertheless, we also present findings that challenge

552     previous hypotheses regarding the environmental drivers of tree species composition. In

553     particular, we suggest that edaphic properties and topography may not always be the most

554     important determinants of floristic composition, and in dynamic floodplain landscapes,

555     disturbance may be a more important driver of tree species composition. Comparing estimates

556     derived from our spectral data with a large dataset of forest plots, we provide compelling

557     evidence for the validity of our approach, not only in classifying broad forest types, but also in

558     describing subtle changes in floristic composition. Finally, our results demonstrate that distance

559     decay relationships are driven by conspecific individuals aggregating at a range of nested scales

560     across landscapes. Reproducing these patterns from theory will require the assumptions of PCP

561     models to be relaxed.

579 **Authors contributions**

580 FCD, GPA and CB conceived the ideas and designed the methodology. GPA collected the

581 imaging spectroscopy data. CB, OLP, RVM, RZG, CAAG, MF, ENHC, TRB, RGV, PVAF, LF,

582 AMM, and RJWB collected the field validation data. FCD, PB and GPA analysed the data. FCD

583 wrote the manuscript with input from GPA, CB and PB. All authors contributed to drafts and

584 gave final approval for publication.

585

586

**References**

588 Adeney, J.M., Christensen, N.L., Vicentini, A. & Cohn-Haft, M. (2016) White-sand Ecosystems
589    in Amazonia. Biotropica, **48**, 7–23.

590 Asner, G.P., Knapp, D.E., Anderson, C.B., Martin, R.E. & Vaughn, N. (2016) Large-scale
591    climatic and geophysical controls on the leaf economics spectrum. Proceedings of the
592    National Academy of Sciences of the United States of America, **113**, E4043-51.

593 Asner, G.P., Knapp, D.E., Boardman, J., Green, R.O., Kennedy-Bowdoin, T., Eastwood, M., …
594    Field, C.B. (2012) Carnegie Airborne Observatory-2: Increasing science data
595    dimensionality via high-fidelity multi-sensor fusion. Remote Sensing of Environment, **124**,
596    454–465.

597 Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J. &
598    Field, C.B. (2007) Carnegie Airborne Observatory: in-flight fusion of hyperspectral
599    imaging and waveform light detection and ranging for three-dimensional studies of
600    ecosystems. Journal of Applied Remote Sensing, **1**, 013536.

601 Assis, R.L., Wittmann, F., Piedade, M.T.F. & Haugaasen, T. (2015) Effects of hydroperiod and
602    substrate properties on tree alpha diversity and composition in Amazonian floodplain
603    forests. Plant Ecology, **216**, 41–54.

604 Baddeley, A. & Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns.
605    Journal of statistical software, **12**, 1–42.

606 Baldeck, C. & Asner, G. (2013) Estimating Vegetation Beta Diversity from Airborne Imaging
607    Spectroscopy and Unsupervised Clustering. Remote Sensing, **5**, 2057–2071.

608 Baldeck, C.A., Tupayachi, R., Sinca, F., Jaramillo, N. & Asner, G.P. (2016) Environmental
609    drivers of tree community turnover in western Amazonian forests. Ecography, **39**, 1089–
610    1099.

611 Baraloto, C., Rabaud, S., Molto, Q., Blanc, L., Fortunel, C., Hérault, B., … Fine, P.V.A. (2011)
612    Disentangling stand and environmental correlates of aboveground biomass in Amazonian
613    forests. Global Change Biology, **17**, 2677–2688.

23

614    Chave, J. & Leigh, E.G. (2002) A spatially explicit neutral model of beta-diversity in tropical
615        forests. Theoretical population biology, **62**, 153–68.

616    Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., …
617        Yamakura, T. (2000) Spatial Patterns in the Distribution of Tropical Tree Species. Science,
618        **288**.

619    Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., … Hubbell, S.P. (2002)
620        Beta-diversity in tropical forest trees. Science (New York, N.Y.), **295**, 666–9.

621    Draper, F.C., Honorio Coronado, E.N., Roucoux, K.H., Lawson, I.T., Pitman, N.C.A., Fine,
622        P.V.A., … Baker, T.R. (2018) Peatland forests are the least diverse tree communities
623        documented in Amazonia, but contribute to high regional beta-diversity. Ecography, **41**, 1–
624        14.

625    Draper, F.C., Roucoux, K.H., Lawson, I.T., Mitchard, E.T.A., Honorio Coronado, E.N.,
626        Lähteenoja, O., Montenegro, L.T., Sandoval, E. V, Zaráte, R. & Baker, T.R. (2014) The
627        distribution and amount of carbon in the largest peatland complex in Amazonia.
628        Environmental Research Letters, **9**.

629    Dumont, J.F., Deza, E. & Garcia, F. (1991) Morphostructural provinces and neotectonics in the
630        Amazonian lowlands of Peru. Journal of South American Earth Sciences, **4**, 373–381.

631    Dumont, J.F., Lamotte, S. & Kahn, F. (1990) Wetland and upland forest ecosystems in Peruvian
632        Amazonia: Plant species diversity in the light of some geological and botanical evidence.
633        Forest Ecology and Management, **33–34**, 125–139.

634    Duque, A., Muller-Landau, H.C., Valencia, R., Cardenas, D., Davies, S., … Vicentini, A. (2017)
635        Insights into regional patterns of Amazonian forest structure, diversity, and dominance from
636        three large terra-firme forest dynamics plots. Biodiversity and Conservation, **26**, 669–686.

637    Duque, Á., Phillips, J.F., von Hildebrand, P., Posada, C.A., Prieto, A., Rudas, A., Suescún, M. &
638        Stevenson, P. (2009) Distance Decay of Tree Species Similarity in Protected Areas on Terra
639        Firme Forests in Colombian Amazonia. Biotropica, **41**, 599–607.

640    Féret, J.-B. & Asner, G.P. (2014a) Microtopographic controls on lowland Amazonian canopy
641        diversity from imaging spectroscopy. Ecological Applications, **24**, 1297–1310.

642    Féret, J.-B. & Asner, G.P. (2014b) Mapping tropical forest canopy diversity using high-fidelity
643        imaging spectroscopy. Ecological Applications, **24**, 1289–1296.

644    Fine, P.V.A., García-Villacorta, R., Pitman, N.C.A., Mesones, I. & Kembel, S.W. (2010) A
645        Floristic Study of the White-Sand Forests of Peru. Annals of the Missouri Botanical
646        Garden, **97**, 283–305.

647    Fortunel, C., Paine, C.E.T., Fine, P.V. a., Kraft, N.J.B. & Baraloto, C. (2014) Environmental
648        factors predict community functional composition in Amazonian forests (ed G De Deyn).
649        Journal of Ecology, **102**, 145–155.

650    Fyllas, N.M., Patiño, S., Baker, T.R., Bielefeld Nardoto, G., Martinelli, L.A., Quesada, C.A., …
651        Lloyd, J. (2009) Basin-wide variations in foliar properties of Amazonian forest: phylogeny,
652        soils and climate. Biogeosciences, **6**, 2677–2708.

653    García-Villacorta, R., Dexter, K.G. & Pennington, T. (2016) Amazonian White-Sand Forests
654        Show Strong Floristic Links with Surrounding Oligotrophic Habitats and the Guiana Shield.
655        Biotropica, **48**, 47–57.

656    García Villacorta, R., Ahuite Reátegui, M. & Olórtegui Zumaeta, M. (2003) Clasificación de
657        bosques sobre arena blanca de la Zona Reservada Allpahuayo-Mishana. Folia Amazonica,
658        **14**, 17–33.

659    Gentry, A.H. (1982) Patterns of Neotropical Plant Species Diversity. Evolutionary Biology, pp.
660        1–84. Springer US, Boston, MA.

661    Guevara, J.E., Damasco, G., Baraloto, C., Fine, P.V.A., Peñuela, M.C., Castilho, C., … ter
662        Steege, H. (2016) Low Phylogenetic Beta Diversity and Geographic Neo-endemism in
663        Amazonian White-sand Forests. Biotropica, **48**, 34–46.

664    Higgins, M.A., Asner, G.P., Perez, E., Elespuru, N., Tuomisto, H., Ruokolainen, K. & Alonso,
665        A. (2012) Use of Landsat and SRTM Data to Detect Broad-Scale Biodiversity Patterns in
666        Northwestern Amazonia. Remote Sensing, **4**, 2401–2418.

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

670  Honorio Coronado, E.N., Baker, T.R., Phillips, O.L., Pitman, N.C.A., Pennington, R.T., Vásquez
671    Martínez, R., … Freitas Alvarado, L. (2009) Multi-scale comparisons of tree composition in
672    Amazonian terra firme forests. Biogeosciences, **6**, 2719–2731.

673  Honorio, E.N., Pennington, T.R., Freitas, L.A., Nebel, G. & Baker, T.R. (2008) Análisis de la
674    composición florística de los bosques de Jenaro Herrera ,. , **15**, 53–60.

675  Hubbell, S.P. (2001) The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32).
676    Princeton University Press.

677  Hubbell, S.P. (2013) Tropical rain forest conservation and the twin challenges of diversity and
678    rarity. Ecology and Evolution, **3**, 3263–3274.

679  Junk, W.J., Piedade, M.T.F., Schöngart, J., Cohn-Haft, M., Adeney, J.M. & Wittmann, F. (2011)
680    A classification of major naturally-occurring Amazonian lowland wetlands. Wetlands, **31**,
681    623–640.

682  Kristiansen, T., Svenning, J.C., Eiserhardt, W.L., Pedersen, D., Brix, H., Munch Kristiansen, S.,
683    … Balslev, H. (2012) Environment versus dispersal in the assembly of western Amazonian
684    palm communities. Journal of Biogeography, **39**, 1318–1332.

685  Kvist, L.P. & Nebel, G. (2001) A review of Peruvian flood plain forests: ecosystems, inhabitants
686    and resource use. Forest Ecology and Management, **150**, 3–26.

687  Levin, S.A. (1992) The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur
688    Award Lecture. Ecology, **73**, 1943–1967.

689  Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M., Baker, T.R. & Phillips, O.L. (2009) Forestplots.
690    net database. University of Leeds, Leeds, UK. Available at: http://www. forestplots.
691    net/(accessed 13 June 2015).

692  Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M. & Phillips, O.L. (2011) ForestPlots.net: a web
693    application and research tool to manage and analyse tropical forest plot data. Journal of
694    Vegetation Science, **22**, 610–613.

695  López Parodi, J. & Freitas, D. (1990) Geographical aspects of forested wetlands in the lower
696    Ucayali, Peruvian Amazonia. Forest Ecology and Management, **33**–**34**, 157–168.

697    Marra, D.M., Chambers, J.Q., Higuchi, N., Trumbore, S.E., Ribeiro, G.H.P.M., dos Santos, J., …
698        Wirth, C. (2014) Large-Scale Wind Disturbances Promote Tree Diversity in a Central
699        Amazon Forest. PLoS ONE, **9**, e103711.

700    Martinez, R.V. & Phillips, O.L. (2000) Allpahuayo: Floristics, Structure, and Dynamics of a
701        High-Diversity Forest in Amazonian Peru. Annals of the Missouri Botanical Garden, **87**,
702        499.

703    May, F., Wiegand, T., Lehmann, S. & Huth, A. (2016) Do abundance distributions and species
704        aggregation correctly predict macroecological biodiversity patterns in tropical forests?
705        Global Ecology and Biogeography, **25**, 575–585.

706    Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., … Green, J.L.
707        (2008) A general framework for the distance-decay of similarity in ecological communities.
708        Ecology letters, **11**, 904–17.

709    Nebel, G., Kvist, L.P., Vanclay, J.K., Christensen, H., Freitas, L. & Ruiz, J. (2001) Structure and
710        floristic composition of flood plain forests in the Peruvian Amazon I. Overstorey. Forest
711        Ecology and Management, **150**, 27–57.

712    Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology.
713        Journal of Biogeography, **26**, 867–878.

714    Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., … Wagner,
715        H. (2013) Package 'vegan.' Community ecology package, version, **2**.

716    Palacios, J., Zarate, R., Torres, G., Denux, J.P., Maco, J.T., Gallardo, G.P., … Cuadros, A.
717        (2016) Mapeo de los bosques tipo varillal utilizando Imagenes de satelite rapideye en la
718        provinica Maynas, Loreto, Peru. Folia Amazónica, **25**, 25–36.

719    Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S. & Wohlgemuth, T. (2002) Quantitative
720        tools for perfecting species lists. Environmetrics, **13**, 121–137.

721    Peacock, J., Baker, T.R., Lewis, S.L., Lopez-Gonzalez, G. & Phillips, O.L. (2007) The
722        RAINFOR database: monitoring forest biomass and dynamics. Journal of Vegetation
723        Science, **18**, 535–542.

724 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay,
725     É. (2012) Scikit-learn: Machine Learning in Python.

726 Phillips, O.L., Baker, T.R., Arroyo, L., Higuchi, N., Killeen, T.J., Laurance, W.F., … Vinceti, B.
727     (2004) Pattern and process in Amazon tree turnover, 1976-2001. Philosophical
728     Transactions of the Royal Society B: Biological Sciences, **359**, 381–407.

729 Phillips, O.L., Martínez, R.V., Vargas, P.N., Monteagudo, A.L., Zans, M.-E.C., Sánchez, W.G.,
730     … Rose, S. (2003a) Efficient plot-based floristic assessment of tropical forests. Journal of
731     Tropical Ecology, **19**, 629–645.

732 Phillips, O.L., Vargas, P.N., Lorenzo, A., Cruz, A.P., Chuspe, M., Sánchez, W.G., … Rose, S.
733     (2003b) Habitat association among Amazonian tree species : a landscape-scale approach. ,
734     757–775.

735 Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., Lafrankie, J. & Ashton, P.S. (2000)
736     Species-area Curves, Spatial Aggregation, and Habitat Specialization in Tropical Forests.
737     Journal of Theoretical Biology, **207**, 81–99.

738 Pos, E.T., Guevara Andino, J.E., Sabatier, D., Molino, J.F., Pitman, N., Mogollón, H., … ter
739     Steege, H. (2014) Are all species necessary to reveal ecologically important patterns?
740     Ecology and Evolution, **4**, 4626–4636.

741 Puhakka, M., Kalliola, R., Rajasilta, M. & Salo, J. (1992) River Types, Site Evolution and
742     Successional Vegetation Patterns in Peruvian Amazonia. Journal of Biogeography, **19**,
743     651–665.

744 Rasanen, M., Nellerf, R.O.N., Saloj, J. & Jungner, H. (1992) Recent and ancient fluvial
745     deposition systems in the Amazonian foreland basin , Peru. , **129**, 293–306.

746 Rocchini, D. (2007a) Effects of spatial and spectral resolution in estimating ecosystem α-
747     diversity by satellite imagery. Remote Sensing of Environment, **111**, 423–434.

748 Rocchini, D. (2007b) Distance decay in spectral space in analysing ecosystem β-diversity.
749     International Journal of Remote Sensing, **28**, 2635–2644.

750 Rocchini, D., Boyd, D.S., Féret, J.-B., Foody, G.M., He, K.S., Lausch, A., … Pettorelli, N.

28

751 (2016) Satellite remote sensing to monitor species diversity: potential and pitfalls (eds A

752 Skidmore and A Chauvenet). Remote Sensing in Ecology and Conservation, **2**, 25–36.

753 Sakschewski, B., von Bloh, W., Boit, A., Poorter, L., Peña-Claros, M., Heinke, J., … Thonicke,

754 K. (2016) Resilience of Amazon forests emerges from plant trait diversity. Nature Climate

755 Change, **6**, 1032–1036.

756 Salo, J., Kalliola, R., Häkkinen, I., Mäkinen, Y., Niemelä, P., Puhakka, M. & Coley, P.D. (1986)

757 River dynamics and the diversity of Amazon lowland forest. Nature, **322**, 254–258.

758 Salovaara, K.J., Thessler, S., Malik, R.N. & Tuomisto, H. (2005) Classification of Amazonian

759 primary rain forest vegetation using Landsat ETM+ satellite imagery. Remote Sensing of

760 Environment, **97**, 39–51.

761 Schwenk, J., Khandelwal, A., Fratkin, M., Kumar, V. & Foufoula-Georgiou, E. (2017) High

762 spatiotemporal resolution of river planform dynamics from Landsat: The RivMAP toolbox

763 and results from the Ucayali River. Earth and Space Science, **4**, 46–75.

764 Socolar, J.B., Gilroy, J.J., Kunin, W.E. & Edwards, D.P. (2016) How Should Beta-Diversity

765 Inform Biodiversity Conservation? Trends in Ecology & Evolution, **31**, 67–80.

766 Soininen, J., McDonald, R. & Hillebrand, H. (2007) The distance decay of similarity in

767 ecological communities. Ecography, **30**, 3–12.

768 Somers, B., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Wright, S.J. & Van De

769 Kerchove, R. (2015) Mesoscale assessment of changes in tropical tree species richness

770 across a bioclimatic gradient in Panama using airborne imaging spectroscopy. Remote

771 Sensing of Environment, **167**, 111–120.

772 Ter Steege, H., Pitman, N.C.A., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., …

773 Castellanos, H. (2006) Continental-scale patterns of canopy tree composition and function

774 across Amazonia. Nature, **443**, 444–447.

775 Ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., …

776 Silman, M.R. (2013) Hyperdominance in the Amazonian tree flora. Science, **342**.

777 Thessler, S. (2008) Remote Sensing of Floristic Patterns in the Lowland Rain Forest Landscape.

778 Tuomisto, H., Poulsen, A.D., Ruokolainen, K., Moran, R.C., Quintana, C., Celi, J. & Cañas, G.

779     (2003a) Linking floristic patterns with soil heterogeneity and satellite imagery in

780     Ecuadorian Amazonia. Ecological Applications, **13**, 352–371.

781 Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003b) Dispersal, environment, and floristic

782     variation of Western Amazonian forests. Science, **299**, 241–244.

783 Wiegand, T., Gunatilleke, S., Gunatilleke, N. & Okuda, T. (2007) Analyzing the spatial structure

784     of a Sri Lankan tree species with multiple scales of clustering. Ecology, **88**, 3088–3102.

785 Wittmann, F., Junk, W.J. & Piedade, M.T.F. (2004) The varzea forests in Amazonia: flooding

786     and the highly dynamic geomorphology interact with natural forest succession. Forest

787     Ecology and Management, **196**, 199–212.

788 Wittmann, F., Schongart, J., Montero, J.C., Motzer, T., Junk, W.J., Piedade, ... & Worbes, M.

789     (2006) Tree species composition and diversity gradients in white-water forests across the

790     Amazon Basin. Journal of Biogeography, **33**, 1334–1347.

791 Zárate, R., Amasifuen, C. & Flores, M. (2006) Floración y Fructificación de plantas leñosas en

792     bosques de arena blanca y de suelo arcilloso en la Amazonía Peruana. Revista Peruana de

793     Biología, **13**.

794

795

796

797 **Table 1**. Summary of the spectral data used to estimate species composition for the three study

798 landscapes

| | **Allpahuayo Mishana** | **Jenaro Herrera** | **Quebrada Braga** |
|---|---|---|---|
| Forest type | White-sand forest | Terra firme forest | Seasonally-flooded forest |
| Total landscape area (ha) | 4540 | 4910 | 3107 |
| Area of forest type (ha) | 794 | 2309 | 2522 |
| No. pairwise comparisons | 315,218 | 2,665,740 | 2,412,585 |

799

800

801

802    **Table 2** Summary of field plot inventory data used to calibrate spectral data at Allpahuayo
803    Mishana (AM) and Jenaro Herrera (JH)

| Plot Type | Large rectangular | Small rectangular | Small circular | Large 'Gentry' | Small 'Gentry' |
|---|---|---|---|---|---|
| Reference | Vasquez & Phillips 2000; Honorio Coronado et al. 2008 | Zarate et al. 2006 | Baldeck et al. 2016 | Baraloto et al. 2011 | (Phillips et al. 2003a) |
| Site | AM & JH | AM | AM | AM & JH | AM |
| No. plots | 15 | 10 | 14 | 4 | 6 |
| Plot area (ha) | 0.5-1.5 | 0.1 | 0.1-0.25 | 0.5 | 0.1 |
| Min. dap (cm) | 5/10 | 2.5 | 10 | 2.5 | 2.5 |
| Mean individuals (per plot) | 663 | 358 | 88 | 242 | 260 |
| Mean identified species (per plot) | 139 | 89 | 33 | 78 | 79 |

804

805

806

807
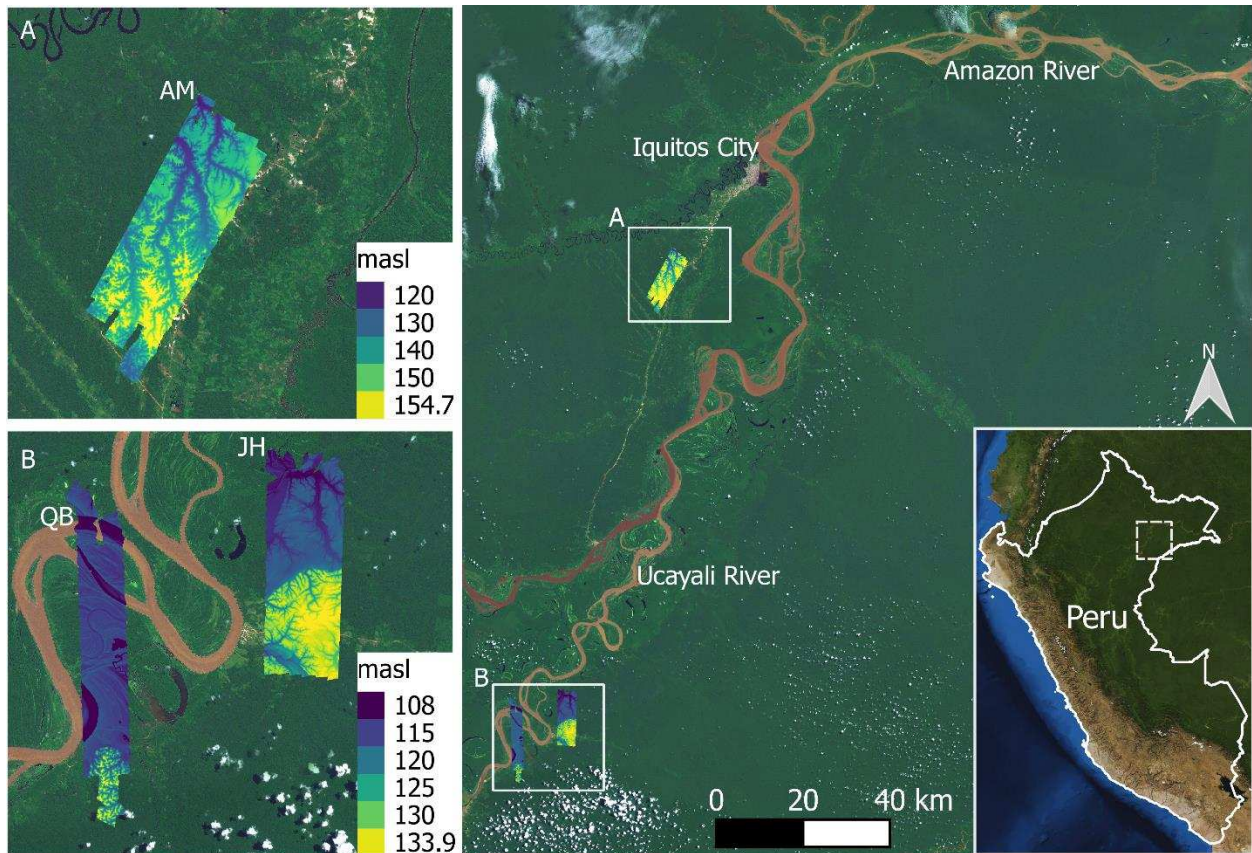
810 **Figure 1.** Maps of the three study landscapes: Allpahuayo Mishana (AM), Quebrada Braga (QB)
811 and Jenaro Herrera (JH). Inset maps A and B show the immediate surroundings of the study
812 landscapes as well as the CAO LiDAR-derived digital terrain models for each landscape. The
813 third inset map shows the wider study region (dashed white box) in the context of Peru.
814

815
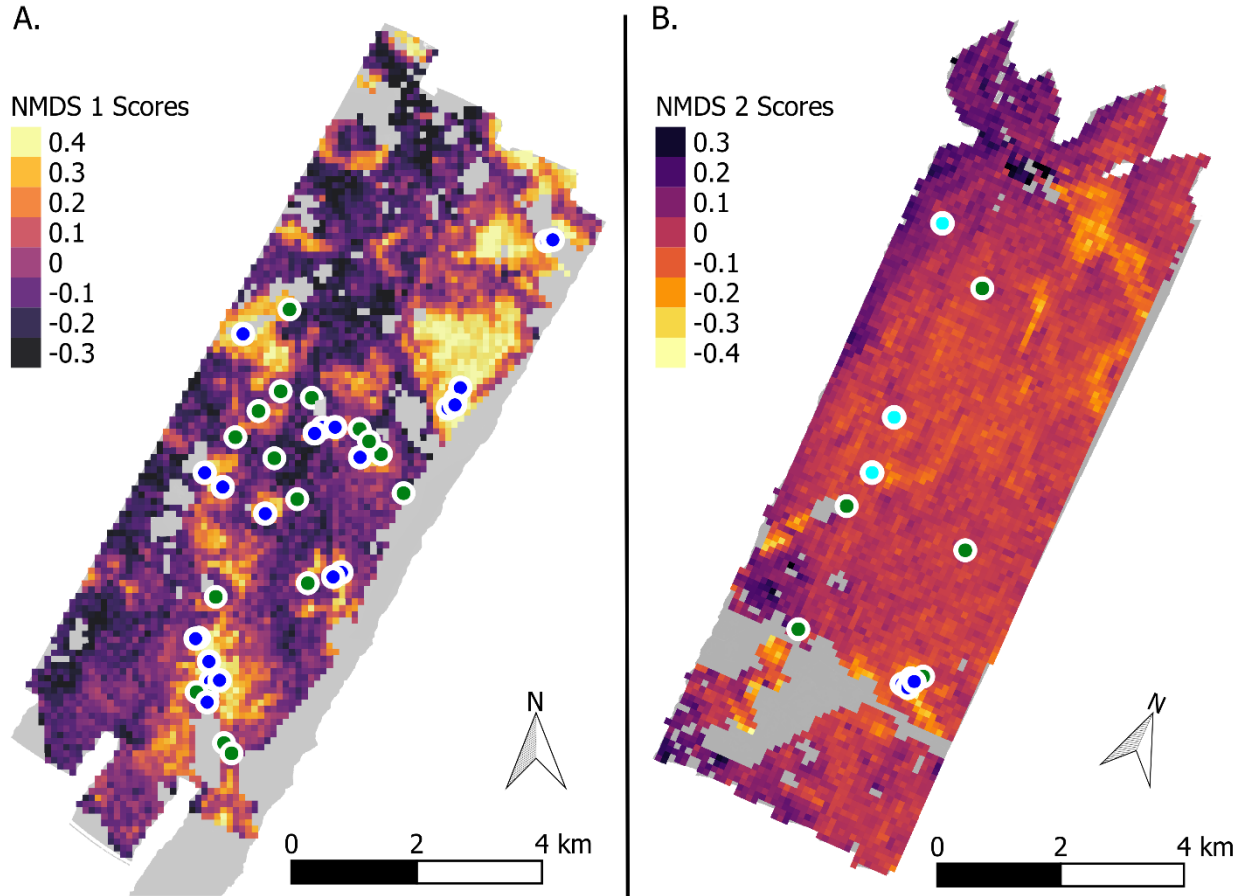


816
817    **Figure 2** Distribution of field plots across the Allpahuayo Mishana landscape (panel A) and the

818    Jenaro Herrera landscape (panel B). Blue circles represent plots in white-sand forest, green

819    circles represent terra firme forest plots and cyan represent palm swamp forest plots. The

820    backdrop of the map shows the first NMDS axis of the estimated species composition of

821    Allpahuayo Mishana and the second NMDS axis of the estimated species composition of Jenaro

822    Herrera, derived from airborne imaging spectroscopy.

823

824

825



826
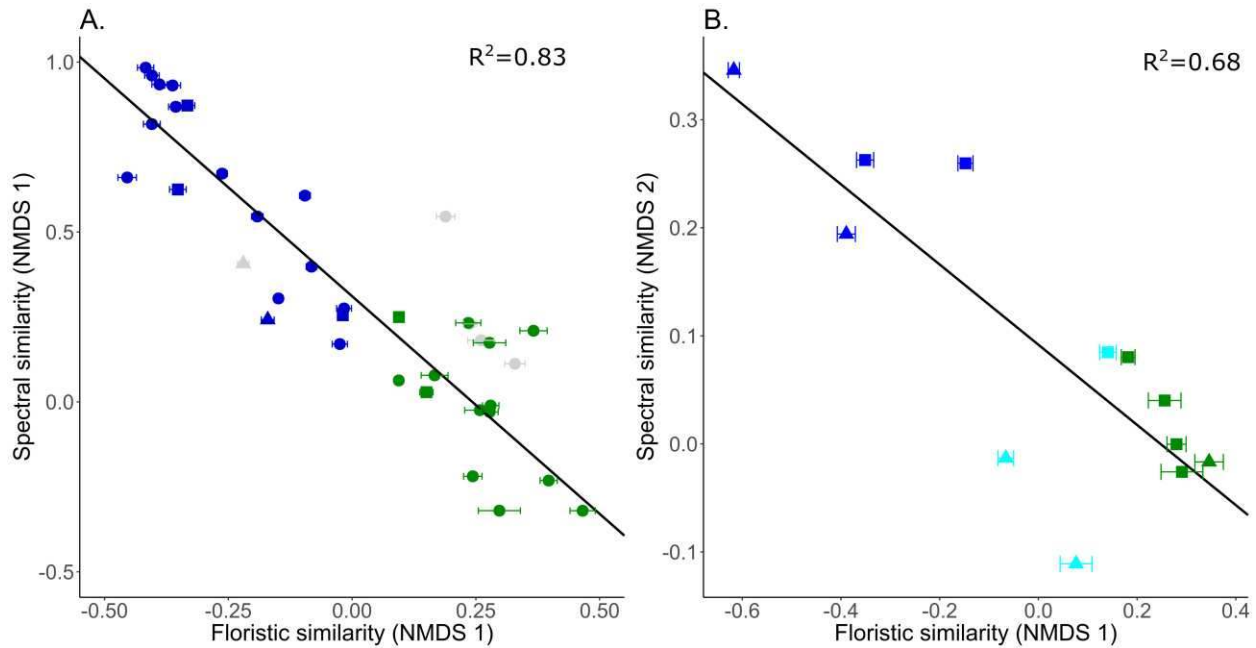
**Figure 3** The relationship between spectrally derived estimates of tree species compositional
turnover (represented by the first axis of the NMDS ordination of spectral species) and measured
tree species compositional turnover (represented by the first axis of the NMDS ordination of tree
species) at Allpahuayo Mishana (panel A.) and Jenaro Herrera (panel B.). Colours represent
different forest types: Dark blue (white-sand forests); green (terra firme clay forests); cyan (palm
swamp forests, grey symbols were those excluded from the analysis are they were < 10 m from a
border between forest types. Error bars signify 95 % confidence intervals around floristic NMDS
axis scores. Symbol shape corresponds to size of forest census plots, square (1 – 1.5 ha), triangle
(0.5 ha), and circle (0.1 – 0.25 ha). Black lines represent linear regressions, both regressions
were highly significant (P ≤ 0.001).

837
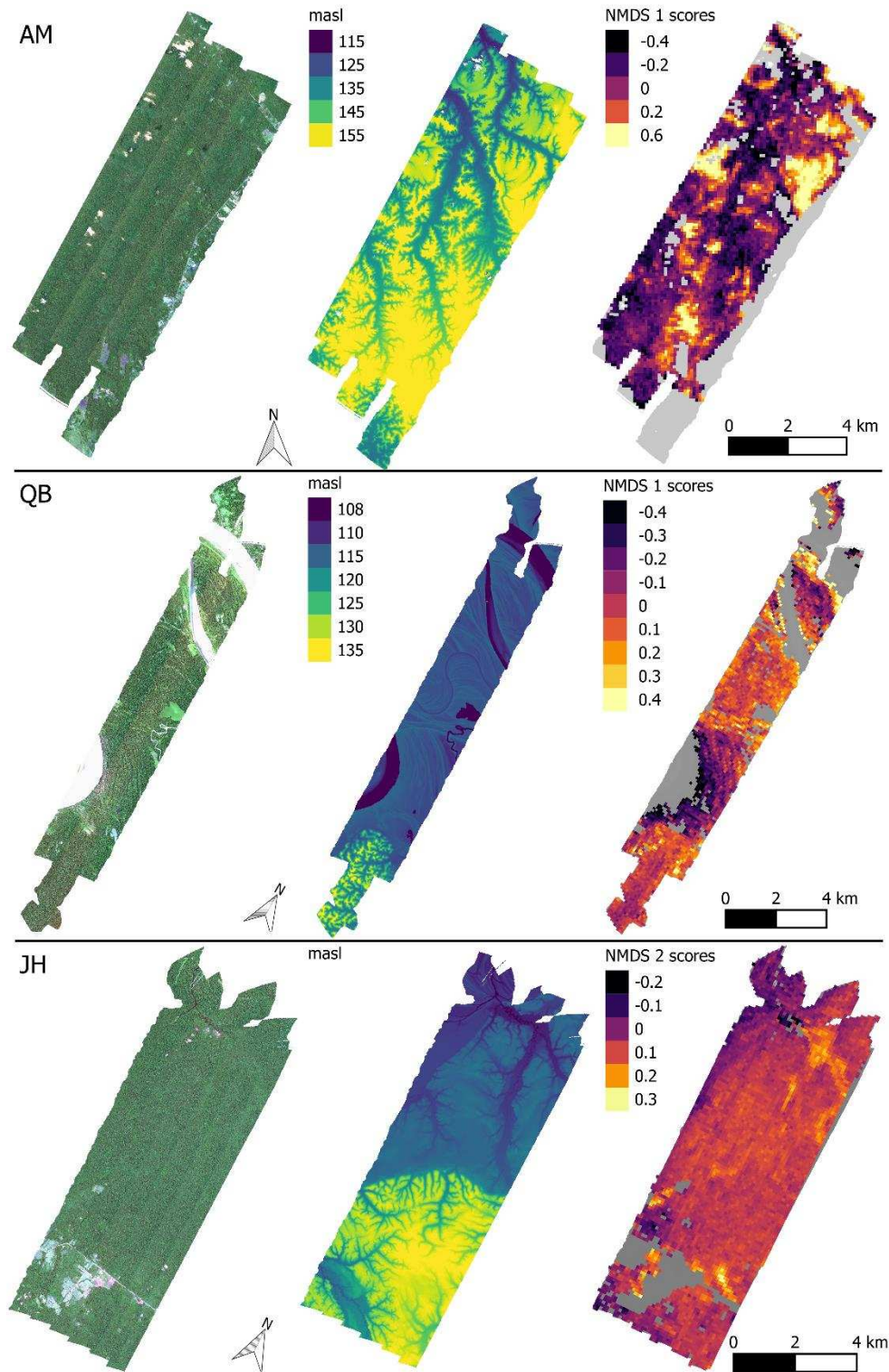
838
**Figure 4** Maps of the three study landscapes, Allpahuayo Mishana (AM), Jenaro Herrera (JH)
and Quebrada Braga (QB), The maps show RGB true colour (column 1), LiDAR-derived
elevation (column 2) and spectrally-derived estimates of tree species composition, summarised
by a single NMDS axis (column 3).

843
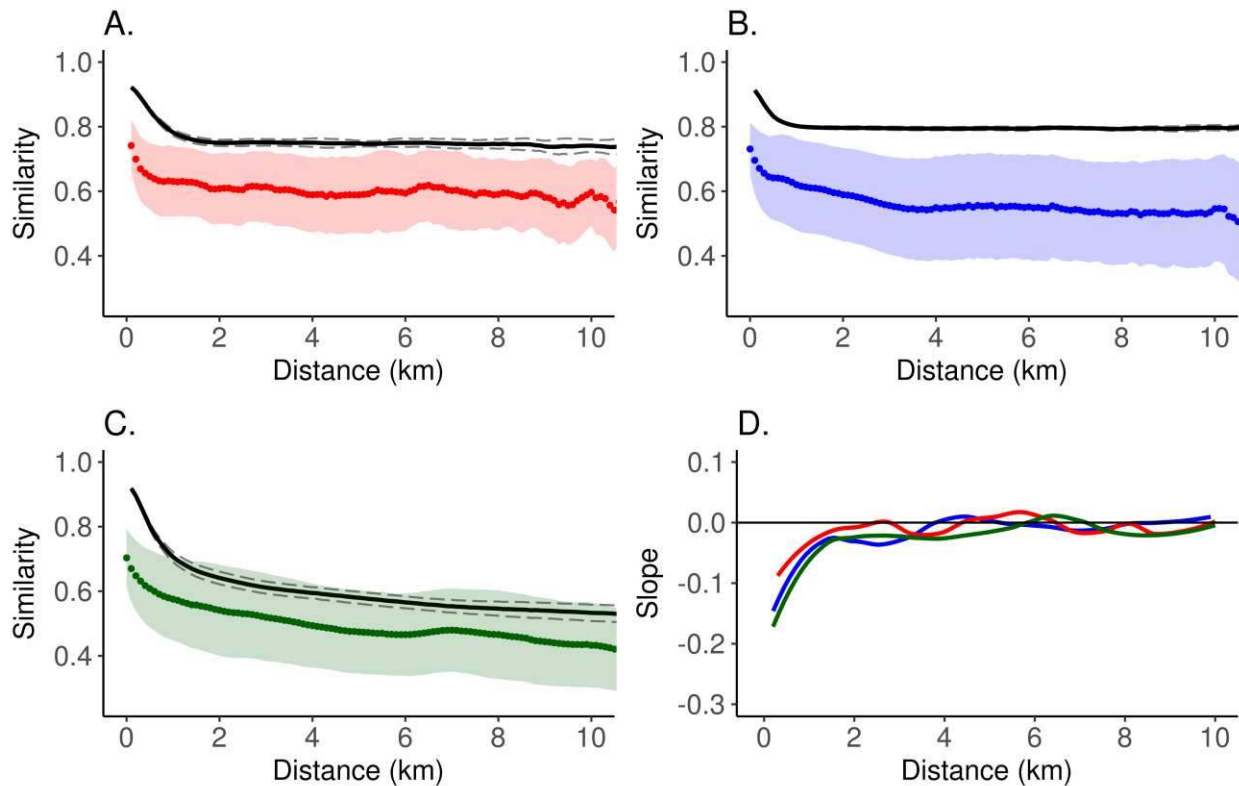


844
845 **Figure 5.** Distance decay relationships in three examples of forest types in the three different
846 landscapes: white-sand forests at Allpahuayo Mishana (panel A); terra firme forest at Jenaro
847 Herrera (panel B); seasonally-flooded forest at Quebrada Braga (Panel C). Points indicate mean
848 Bray-Curtis indices of similarity every 100 meters, and shaded areas are the standard deviations
849 surrounding each 100m point. Panel D shows the loess smoothed line (span=0.35) through the
850 first order derivative, calculated every 100m at each site. Colours correspond to different
851 landscapes/forest types: red = Allpahuayo Mishana white-sand, blue = Jenaro Herrera terra firme
852 clay, green = Quebrada Braga seasonally-flooded. Solid black lines indicate the mean PCP
853 theoretical predicted distance decays, and dashed black lines the standard deviations surrounding
854 these means.
855