This is the peer reviewed version of the following article: Marselis, SM, Abernethy, K, Alonso, A, et al. Evaluating the potential of fullwaveform lidar for mapping pan-tropical tree species richness. *Global Ecol Biogeogr*. 2020; 29: 1799-1816, which has been published in final form at https://doi.org/10.1111/geb.13158. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for self-archiving.

SI1

Here we provide information on the characteristics of each field and accompanying lidar dataset.

lop (Lopé, Gabon)

The *lop* field dataset in Gabon was collected as part of the AfriSAR campaign. The study site is located in Central Gabon in the northern section of the Lopé National Park. The field dataset was collected in 2016 and published in Labrière et al. (2018). The monthly minima and maxima temperature ranges between 20-23 °C and 26-33 °C, respectively. Mean annual precipitation is ~1440 mm/year. The dataset consists of 12 field plots, distributed over five vegetation types: Savanna (3), Colonizing forest (3), monodominant Okoumé forest (3), young Marantaceae forest (2) and mixed Marantaceae forest (1). We only used data from 11 plots as one savanna site had no trees with DBH ≥ 10 cm. The three colonizing forest plots are 0.5 ha (100x50 m). All other plots are 1.0 ha (100x100 m). Each plot was stem mapped and all trees with DBH ≥ 5 cm were recorded for colonizing forests and DBH ≥ 10 cm in all other forest types (Labrière et al., 2018). Because of the presence of stem maps all 1.0 ha plots were used at all three spatial resolutions in this study, while the 0.5 ha plots were only used at the 0.25 and 0.0625 ha resolutions. All plots had overlapping airborne lidar data collected in 2016 with the NASA Land Vegetation and Ice Sensor (LVIS) instrument (Tang et al., 2018).

mon (Mondah, Gabon)

The *mon* field dataset in Gabon was collected as part of the AfriSAR campaign in 2016 and this data is publicly accessible through the NASA Distributed Active Archive Center (Fatoyinbo et al., 2018). The mean annual temperature in this study site is ~25°C, with a precipitation ranging between 3000-3500 mm/year. The forests across the different field plots have experienced different degrees of disturbance, ranging from slightly disturbed to highly disturbed, given their proximity to Gabon’s capital Libreville

(Walters et al., 2016). The data consists of 15 1 ha field plots, for which each tree with DBH ≥ 5 cm was located, identified and measured. Only 12 of the field plots had overlapping lidar data, collected in 2016 with NASA Land Vegetation and Ice Sensor (LVIS) instrument (Tang et al., 2018). Given the presence of stem maps, the plots could be used at all three spatial resolution by creating subplots and using the tree location to assign the trees to the respective subplots. Of the 12 plots only 10 had a species identification quality high enough (> 80% of the trees identified to at least the genus level) to be included in the analyses at the 1.0 ha resolution, 11 of the original plots were included in the 0.25 ha resolution analyses and 12 of the plots were included at the 0.0625 ha resolution analyses. This was possible as at the smaller plot sizes some of the subplots within the original plot did have a sufficiently high percentage of identified trees.

mab (Mabounie, Gabon)

The *mab* study site is located in west-central Gabon. This study site has experienced a lower degree of forest degradation than the *mon* study site. The mean annual temperature lays at 25°C and mean annual precipitation at roughly 2050 mm/year[1]. Wet and dry seasons occur. The study site consists of 12 field plots of 1 ha collected in 2012. For each of the plots, all trees with DBH ≥ 10 cm were assigned within 20x20 m subplots and identified to the species level. 10 of the original plots had overlapping airborne lidar data collected in 2016 with NASA Land Vegetation and Ice Sensor (LVIS) instrument (Tang et al., 2018). The plots from this study site were only used at the 1.0 ha plot size because the native subplot size was unsuitable for creating subplots of 25x25 (0.0626 ha) or 50x50 m (0.25 ha).

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

rab (Rabi, Gabon)

The *rab* (Rabi) study site in Gabon is located in southwestern Gabon. The region experiences an annual rainfall of approximately 2300 mm/year and a mean annual temperature between 25-28 °C (Labrière et al., 2018). The data was collected through a partnership of the Smithsonian Institute and Gabon's National Center for Scientific and Technological Research (CENAREST). All tree measurements were collected over a two year time period between June 2010 and June 2012 (Memiaghe et al., 2016). A 25 ha plot was stem mapped and for each tree with DBH ≥ 1 cm, the species, size and location were recorded (Memiaghe et al., 2016). Due to the presence of the stem map, this study site could be used at all spatial resolutions. For the regional scale analysis, we included only the four 1 ha plots at each corner of the square 500 x 500 m plot to reduce the mixing of the local and regional relationships between canopy structure and tree species richness. The plot has overlapping airborne lidar data collected in 2016 with NASA's Land Vegetation and Ice Sensor (LVIS) instrument (Tang et al., 2018).

mal (Malebo, DRC)

The *mal* dataset was collected in 2011 in the western part of the DRC. The mean annual temperature lays around 24.9 °C and the mean annual precipitation is measured to be 1596 mm/year[2]. The dataset consists of 32 1-ha field plots in which each tree of DBH ≥ 10 cm is located inside a 50x50 m subplot. Therefore, this dataset was only used for analyses at the 0.25 and 1.0 ha resolutions. Lidar data was collected in 2014 coincident with 21 of the plots, which all had a high species identification quality and were included in this paper.

yan (Yangambi, Central Congo Basin, DRC)

The *yan* dataset was collected in the UNESCO Man and Biosphere reserve in Yangambi in the Democratic Republic of the Congo and published by Kearsley *et al.* (2013). The average annual precipitation is 1762

[2] http://www.worldclim.org/

mm/year with an average temperature of ~25 °C[2]. The forest plots cover different types of forest ranging from young secondary forest to old growth moist evergreen rainforest (monodominant Gilberiodendron forest and mixed forest). In this study we included only 9 of the plots, covering the whole variety of forest structure, for which we had coincident airborne lidar data available. The permanent sampling plots were installed in 2012 to establish a baseline reference data on carbon stocks and biodiversity for this region. Each plot was stem mapped and monitoring of these plots is ongoing. Species identification was performed in the field by two independent botanists. Additionally, for the plots in young regrowth and old-growth forest, at least one herbarium specimen and one silicagel dried leaf sample was collected for each taxon. Vouchers were compared with reference specimens kept by the Botanic Garden Meise (Belgium) for both morphologic and genetic characteristics (barcodes generated using rbcL and matK sequences as recommended by the CBOL Plant Working Group 2009). (Kearsley et al., 2013). The nine 1 ha stem mapped plots for which all trees with a DBH ≥ 10 cm were located, identified and measured were included in this analysis at all three spatial resolutions. The coincident lidar data was collected in 2014 (Xu et al., 2017).

s11 (SFX 2011, Brazil)

In 2011 a dataset consisting of nine 40 x 40 m field plots was collected in the São Félix do Xingu region located in the State of Pará in Brazil[3]. The region experiences a mean annual temperature of 25.1˚C and a mean annual precipitation around 1972 mm/year[4]. The plots are located in a linear fashion, originally meant to cover areas where ICESat-1 waveforms were collected, along a north-south transect. The type of forest ranges from old-growth forest to highly disturbed, through (multiple) fire(s) and/or logging. The data collection was carried out through the Paisagens Sustentaveis (Sustainable Landscapes)

[3] https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/c141496b-1fe3-40aa-926b-d2c88903df97
[4] http://www.worldclim.org/

program funded through the US Agency for International Development and the US Department of State. Each tree with a DBH ≥ 10 cm was measured, identified and located within each field plot, allowing this dataset to be used only at the 0.0625 ha resolution. Eight of the nine plots were used in this paper, as these met species identification quality requirements. The US Forest Service in collaboration with the Brazilian Agricultural Research Corporation (Embrapa) collected coincident lidar data over all field plots in 2012 with an ALTM 3100 scanner, and a resulting point density of 30.1 points/m^2. Both datasets are available through the Sustainable Landscapes data portal[5]. Field identifications were done by parataxonomists. In general the identifications were not checked by professional botanists nor were voucher specimens taken in most cases, thus it was impossible to evaluate the quality of the species identifications as they were collected primarily to estimate wood density for biomass estimates.

s12 (SFX 2012, Brazil)

The *s12* dataset was a follow-up field data collection on the *s11* dataset. This dataset has the same characteristics as the *s11* dataset, but it was collected in 2012 about 100 km west of the original dataset, but with the same spatial distribution of the field plots along ICESat-1 data collection lines[6]. The dataset consists of 22 field plots of 40x40 m, including all trees with a DBH ≥ 10 cm. These plots were also only used at the 0.0625 ha resolution due to their native plot size. 21 plots were used in the analysis, as there was one plot that did not meet the 80% identified to genus-level requirement.

tam (Tambopata, Peru)

The *tam* field dataset was collected in Tambopata, lowland Amazonian rainforest in Peru on June 7th, 2009. The annual rainfall in this region is ~2250 mm/year and the mean annual temperature is

[5] https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/a3bb4f79-ef32-4b28-8295-2d963a1044e5
[6] https://www.paisagenslidar.cnptia.embrapa.br/geonetwork/srv/por/catalog.search#/metadata/43b6c844-fc8d-42c4-9e54-0946b7e101e8

measured at 25.4˚C[7]. The study site lays at the boundary of a nature reserve were sustainable use of forest resources is granted. Data were originally collected across seven plots, but one of these was located in a swamp area and not included in this study. The plots are 100 x 100 m (1 ha) and each tree with a DBH ≥ 10 cm was located, identified and measured. The airborne lidar data was acquired in 2009 with a Leica ALS50 II System. This dataset has a point density of 2.1 points/m^2 (Boyd et al., 2013).

bci (Barro Colorado Island, Panama)

The 50 ha permanent plot in Barro Colorado Island, Panama, is by far the most well-studied site included in this study. This site is the only one for which analyses on the relationship between canopy structure and tree species diversity have been carried out previously by other scientists (Fricker et al., 2015; Wolf et al., 2012). The study site consists of a 500 x 1000 m plot on the Barro Colorado Island with a mean annual precipitation of ~2580 mm/year and a mean annual temperature is 26.5˚C[8]. The plot was first established in 1980 and has since been measured every five years. All these data are available through the ForestGEO data portal[9]. In this analysis we used the data from the 2010 survey. Each tree with a DBH ≥ 1 cm is marked, located, identified and measured. The data from the *bci* study site were used at all spatial resolutions because of the availability of a stem map, but we only used 6 of the original 50 ha for the regional and pan-tropical analysis. The lidar data were collected over BCI in 2012 (Lobo and Dalling, 2013).

lsv (La Selva, Costa Rica)

The *lsv* dataset was collected as part of the CARBONO project at the La Selva Research Station in Costa Rica (Clark and Clark, 2000). Mean annual temperature is 25.9˚C, mean annual precipitation approximately 4035 mm/year[8]. The study site consists of well-preserved old-growth tropical forest. The

[7] http://www.worldclim.org/
[8] http://www.worldclim.org/
[9] https://forestgeo.si.edu/sites/neotropics/barro-colorado-island

field study project focused on the long-term monitoring of tropical rainforest productivity at the landscape scale[10]. Field data has been collected in 18 50 x 100 (0.5 ha) plots for twenty-one consecutive years between 1997 and 2017. In this study we used the data from the 2009 census as this one was the closest in date to the lidar data collection. All trees were identified and stem-mapped and the data were used in the analyses at the 0.25 and 0.0625 ha resolution. The lidar data were collected with the LVIS instrument in 2009 coincident with all 18 field plots.

cha (La Selva 2, Costa Rica)

The *cha* dataset was also collected in the same region as the *lsv* dataset, on the east side of the La Selva biological station in Costa Rica. The dataset consists of two 0.5 ha plots and one 1 ha plot. All trees with a DBH ≥ 5 cm were located, identified and measured. The vegetation sampled covers regrowth forest of ages between 23 and 31 years old. The plots were used at three or two spatial resolutions, depending on the size of the original plot. The same lidar dataset collected by LVIS in 2009 coincided with these field plots.

rob (Robson Creek, Australia)

The *rob* study site is comprised of a 500 x 500 m (25 ha) plot located in Northern Queensland, Australia (Bradford et al., 2014). Mean annual temperature is ~20.6 °C. The plot is considered a so-called supersite and was established as part of the Terrestrial Ecosystem Research Network (TERN). The data are available through the Australian Supersite Network portal[11]. Mean annual precipitation in the region is measured at 1597 mm/year (Bradford et al., 2014). All trees with a DBH ≥ 10 cm were located, identified and measured, allowing the dataset to be used for the analyses at the 0.0625, 0.25 and 1.0 ha resolutions. Lidar data were collected with a Riegl VZQ560 airborne lidar over the study site in 2012 by

[10] https://tropicalstudies.org/carbono-project/

[11] https://supersites.tern.org.au

the AusCover facility of TERN. For the regional and pan-tropical analysis only the four 1 ha corner plots were used.

dan (Danum Valley, Malaysia)

The *dan* dataset comprises a 50 ha (500 x 1000) plot in the Danum Valley conservation area in Malaysia. The data is available through the ForestGeo website[12]. Mean annual temperature at the study site is 26̊ C, mean annual precipitation approximately 2890 mm/year[13]. The plot was established in 2011 and all trees of DBH ≥ 1 cm were located and measured. However, the tree species identification of this field plot is still ongoing at the time of the analyses in this study and therefore this plot was not used in the local scale analysis, even though it comprised a large continuous area, because not every adjacent 1 ha plot had the required level of species identification (> 80% of the trees identified to the genus level) as species identification was still ongoing at the time of these analyses. The six corner 1 ha plots and the middle 1 ha plots along the long side of the plot were considered for the regional and global analysis. Of these six plots, two of them had the required tree species identification quality at the 0.25 and 1 ha resolutions and were included in the regional and pan-tropical analysis. At the 0.0625 ha resolution we were able to include at least one subplot of each of the original selected six 1 ha plots. The lidar data collection was funded through the National Environmental Research Council (NERC) and collected in 2014 with a Leica ALS50-II lidar system. The lidar dataset had a point density of 7.3 points/m^2.

sep (Sepilok, Borneo)

The *sep* dataset consists of nine 4 ha plots (200 x 200 m) located in the Sepilok Forest Reserve north-east of Borneo. The mean annual precipitation in this region is around 2929 mm and mean annual temperature between 26.7-27.7 °C (DeWalt et al., 2006). These field data were collected in 2014. From

[12] https://forestgeo.si.edu/sites/asia/danum-valley
[13] http://www.worldclim.org/

each original 4 ha field plot, one 1 ha plot (or one 0.25 ha plot, depending on the resolution of the analysis) was selected at random during the Monte Carlo simulations to avoid the confusion of the local and regional variation in canopy structure and tree species diversity. The *sep* lidar dataset covered all nine field plots, hence all were included in the analyses.

Data processing

The data processing resulted in a variable number of plots at each resolution for each dataset. These numbers are represented in Table SI1-1.

Table SI1-1: Number of plots included at each spatial resolution for each dataset with a percentage of trees identified up to genus level > 80%. For study sites with a plot size > 1 ha, a subset of the original plot was used at each of the analyses as described in the text above and indicated with the two numbers for one dataset: the first number indicating the total number of available plots at the given resolution and the second number the total number of plots used at that resolution. The sep *plot in Borneo was a special case for which a nine plots of 4 ha each were available, which were all included in the analysis but one 1 ha plot from the original 4 was selected in each Monte Carlo simulation.*

COUNTRY	PROJECT NAME	NO. NATIVE PLOTS	TOTAL AREA (HA)	# OF 1.0 HA PLOTS	# OF 0.25 HA PLOTS	# OF 0.0625 HA PLOTS
Oceania						
Australia	*rob*	1	25	25/4	100/16	400/64
South-East Asia						
Malaysia	*dan*	1	50	20/2	78/8	321/36
Borneo	*sep*	9	36	36/36	144/144	-
Africa						
DRC	*mal*	21	21	21	62	-
DRC	*yan*	9	9	9	35	140
Gabon	*rab*	1	25	25/4	100/16	399/64
Gabon	*lop*	11	9.5	8	37	140
Gabon	*mon*	12	12	10	41	137
Gabon	*mab*	10	10	10	-	-
South America						
Peru	*tam*	6	6	6	24	96
Brazil	*s11*	8	1.44	-	-	8
Brazil	*s12*	21	3.36	-	-	21
Central America						
Costa Rica	*lsv*	18	9	-	36	144
Costa Rica	*cha*	3	2	1	8	32
Panama	*bci*	1	50	50/6	179/24	726/96

Simulated GEDI waveforms were created from each airborne lidar dataset using the GEDI waveform simulator (Hancock et al., 2019). The simulator takes discrete return airborne laser scanning data, or LVIS lidar waveforms, and makes them look like GEDI waveforms. The waveforms were processed using the gediMetric software, which extracts metrics of interest from the simulated waveforms. In this study we used the rhReal98 metric to reflect canopy height, and the hiLAI profile to represent the PAI profile as this method provides a profile is most similar to profiles derived from the LVIS waveforms verified and used in (Marselis et al., 2019, 2018).

Environmental variables

In this study we aimed to only look at the relationship between canopy structure and tree species richness. Our field plots were selected to cover a large range in canopy structure and species richness by including forest plots across different gradients of succession, degradation and maturity. In this study we did not look at the relationship between environmental variables and tree species richness, even though such variables are known to be related to tree species diversity, as the plots used here were not chosen to cover an environmental gradient. Figure S1-1 shows the relationship between species richness and mean annual temperature and precipitation, illustrating that all plots within the same study site have highly similar mean annual temperature and precipitation as they are all closely located together. Hence using 15 different combinations of temperature and precipitation in a combined environmental – canopy structure model to explain tree species richness would provide highly erroneous results and thus we have refrained from doing so in this study.

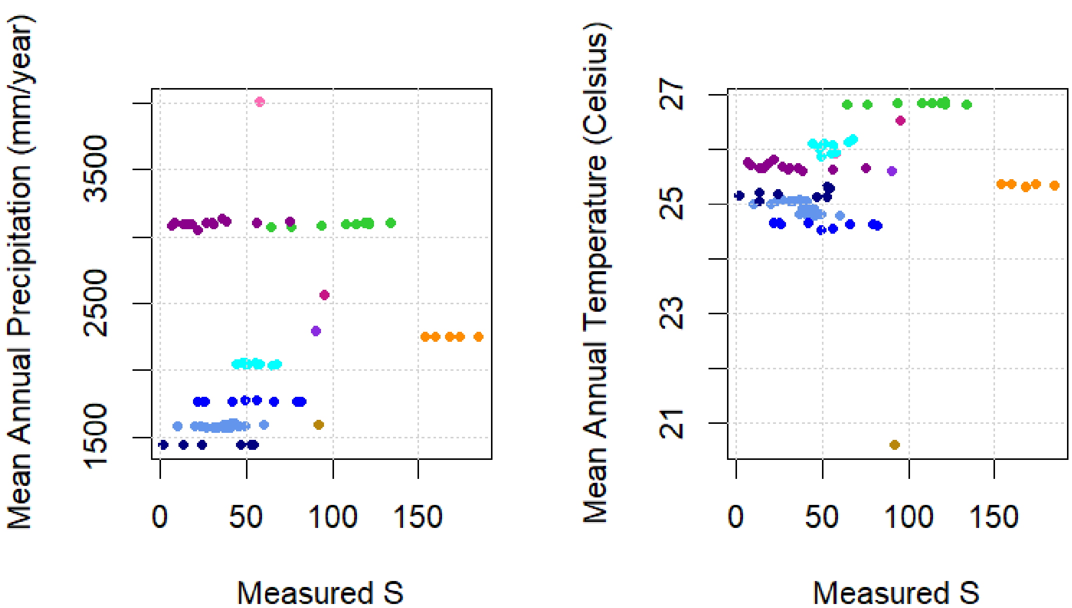

Figure S1-1: Relation between Measured S (tree species richness at the 1.0 ha plot size) and mean annual precipitation and temperature. Our plots demonstrate a very limited range of environmental variation given the close location of plots in each study site.

References

Boyd, D.S., Hill, R.A., Hopkinson, C., Baker, T.R., 2013. Landscape-scale forest disturbance regimes in southern Peruvian Amazonia. Ecol. Appl. 23, 1588–1602. https://doi.org/10.1890/12-0371.1

Bradford, M.G., Metcalfe, D.J., Ford, A., Liddell, M.J., McKeown, A., 2014. Floristics, stand structure and aboveground biomass of a 25-ha rainforest plot in the wet tropics of Australia. J. Trop. For. Sci. 26, 543–553.

Clark, D.B., Clark, D.A., 2000. Landscape-scale variation in forest structure and biomass in a tropical rain forest. For. Ecol. Manage. 137, 185–198.

DeWalt, S.J., Ickes, K., Nilus, R., Harms, K.E., Burslem, D.F.R.P., 2006. Liana habitat associations and community structure in a Bornean lowland tropical forest. Plant Ecol. 186, 203–216. https://doi.org/10.1007/s11258-006-9123-6

Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57–68. https://doi.org/10.1007/BF00038687

Fatoyinbo, T.E., Saatchi, S.S., Armston, J.D., Poulsen, J., Marselis, S.M., Pinto, N., White, L.J.T., Jeffery, K.J., 2018. AfriSAR: Mondah Forest Tree Species, Biophysical, and Biomass Data, Gabon, 2016. ORNL DAAC, Oak Ridge, Tenessee, USA. https://doi.org/https://doi.org/10.3334/ORNLDAAC/1580

Fricker, G.A., Wolf, J. a., Saatchi, S.S., Gillespie, T.W., 2015. Predicting spatial variations of tree species richness in tropical forests from high resolution remote sensing. Ecol. Appl. 25, 150218095111005. https://doi.org/10.1890/14-1593.1

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., Dubayah, R., 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. Earth Sp. Sci. 6, 294–310. https://doi.org/10.1029/2018EA000506

Kearsley, E., De Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., Huygens, D., Kebede, Y., Defourny, P., Bogaert, J., Beeckman, H., Steppe, K., Boeckx, P., Verbeeck, H., 2013. Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. Nat. Commun. 4. https://doi.org/10.1038/ncomms3269

Labrière, N., Tao, S., Chave, J., Scipal, K., Le Toan, T., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Casal, T., others, 2018. In Situ Reference Datasets From the TropiSAR and AfriSAR Campaigns in Support of Upcoming Spaceborne Biomass Missions. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–11.

Lobo, E., Dalling, J.W., 2013. Effects of topography, soil type and forest age on the frequency and size distribution of canopy gap disturbances in a tropical forest. Biogeosciences 10, 6769–6781. https://doi.org/10.5194/bg-10-6769-2013

Marselis, S.M., Tang, H., Armston, J., Abernethy, K., Alonso, A., Barbier, N., Bissiengou, P., Jeffery, K., Kenfack, D., Labrière, N., others, 2019. Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. Environ. Res. Lett. 14. https://doi.org/https://doi.org/10.1088/1748-9326/ab2dcd

Marselis, S.M., Tang, H., Armston, J.D., Calders, K., Labrière, N., Dubayah, R., 2018. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. Remote Sens. Environ. 216, 626–634.

Memiaghe, H.R., Lutz, J.A., Korte, L., Alonso, A., Kenfack, D., 2016. Ecological Importance of Small-Diameter Trees to the Structure, Diversity and Biomass of a Tropical Evergreen Forest at Rabi, Gabon. PLoS One 11. https://doi.org/10.1371/journal.pone.0154988

Tang, H., Armston, J.D., Hancock, S., Hofton, M., Blair, J.B., Dubayah, R.O., 2018. AfriSAR: Canopy Cover and Vertical Profile Metrics Derived from LVIS, Gabon, 2016. ORNL DAAC, Oak Ridge, Tenessee, USA.

Walters, G., Ngagnia Ndjabounda, E., Ikabanga, D., Biteau, J.P., Hymas, O., White, L.J.T., Ndong Obiang, A.-M., Ndong Ondo, P., Jeffery, K.J., Lachenaud, O., Stévart, T., 2016. Peri-urban conservation in the Mondah forest of Libreville, Gabon: Red List assessments of endemic plant species, and avoiding protected area downsizing. Oryx 50, 419–430. https://doi.org/10.1017/S0030605315000204

Wolf, J.A., Fricker, G.A., Meyer, V., Hubbell, S.P., Gillespie, T.W., Saatchi, S.S., 2012. Plant species richness is associated with canopy height and topography in a neotropical forest. Remote Sens. 4, 4010–4021. https://doi.org/10.3390/rs4124010

Xu, L., Saatchi, S.S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., Bastin, J.F., Banks, N., Boeckx, P., Verbeeck, H., Lewis, S.L., Muanza, E.T., Bongwele, E., Kayembe, F., Mbenza, D., Kalau, L., Mukendi, F., Ilunga, F., Ebuta, D., 2017. Spatial Distribution of Carbon Stored in Forests of the Democratic Republic of Congo. Sci. Rep. 7, 15030. https://doi.org/10.1038/s41598-017-15050-z

SI2

Example plots of observed vs. predicted tree species richness are provided here to illustrate the meaning of low R^2 values accompanied by low RMSD%. This model performance indicates that at the local scale in *bci* the species richness predictions (x-axis) are around the average observed S (y-axis), but there is low sensitivity to the local variation in tree species richness within the plot (Figure SI2-1).

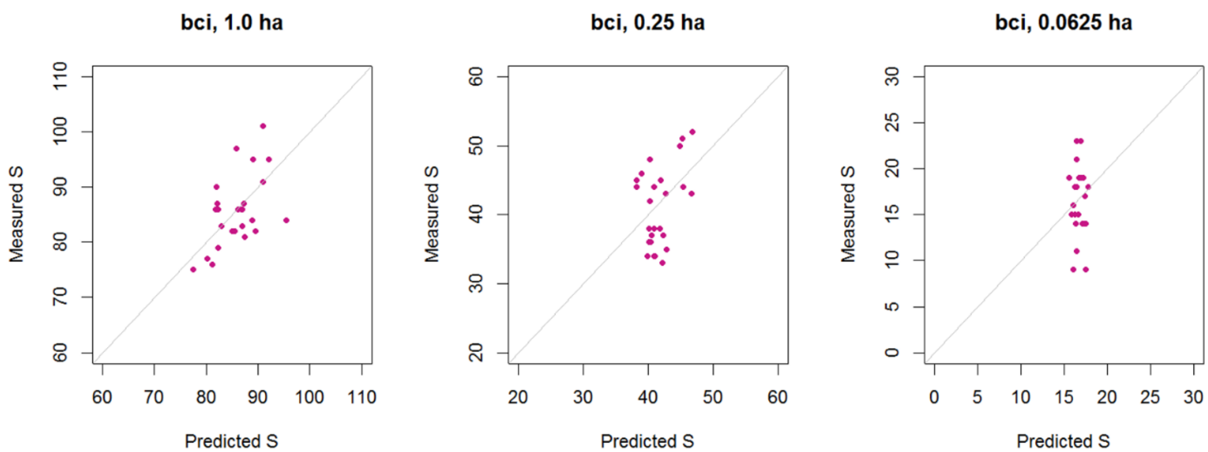

Figure SI2-1: Predicted vs. Measured species richness from local scale predictions in *bci*.

SI3

We computed the ecological distance between each of the plots in relation to the spatial distance between these plots. We used the method proposed by Bray-Curtis to calculate this distance (Faith et al., 1987). This method takes a table with the species in the columns, the plots in the rows and the number of trees of each species in each plot as content and computes a measure of similarity between the compositions expressed as the semi-variance ranging between 0-1. The relationships are shown for each study site at the 0.25 ha resolution (or 0.16 ha in case the 0.25 ha plots were not available due to the native plot size) in figure SI3-1 to SI3-15.

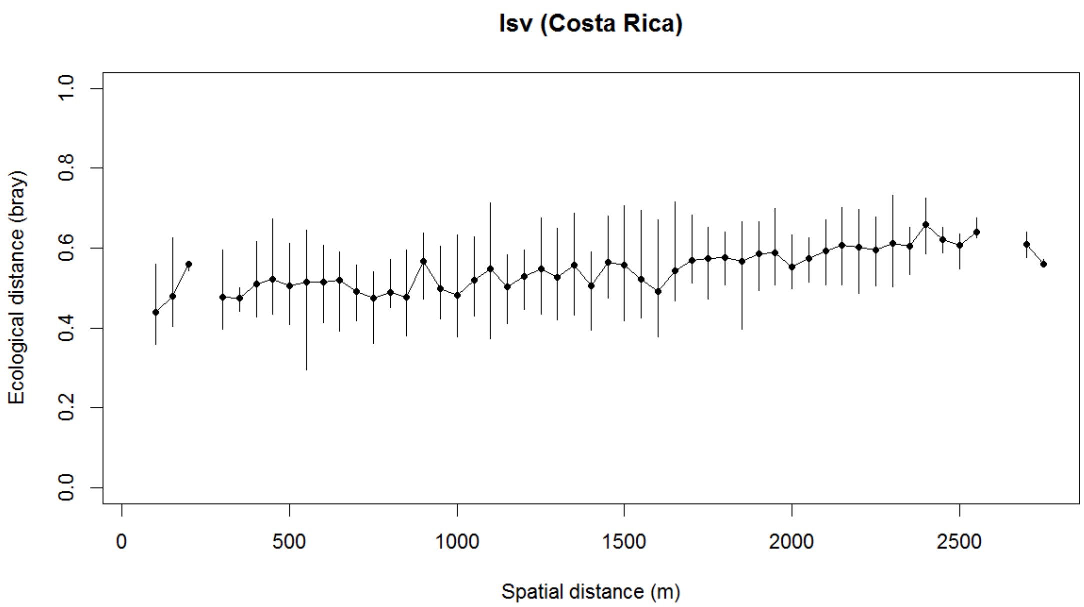

Figure SI3-1: Ecological distance vs. spatial distance in Isv, Costa Rica

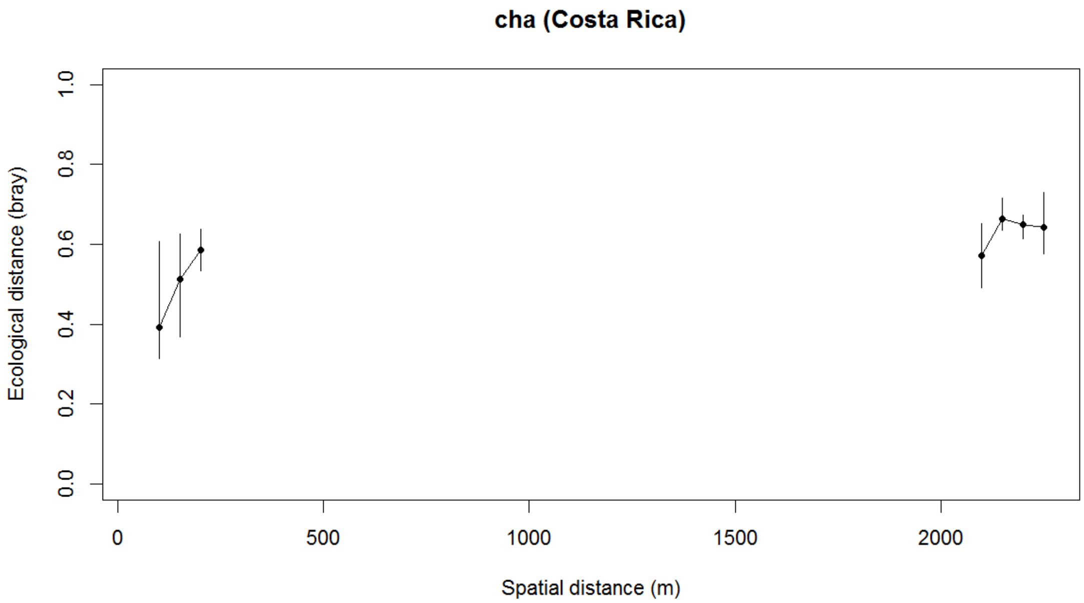

Figure SI3-2: Ecological distance vs. spatial distance in cha, Costa Rica

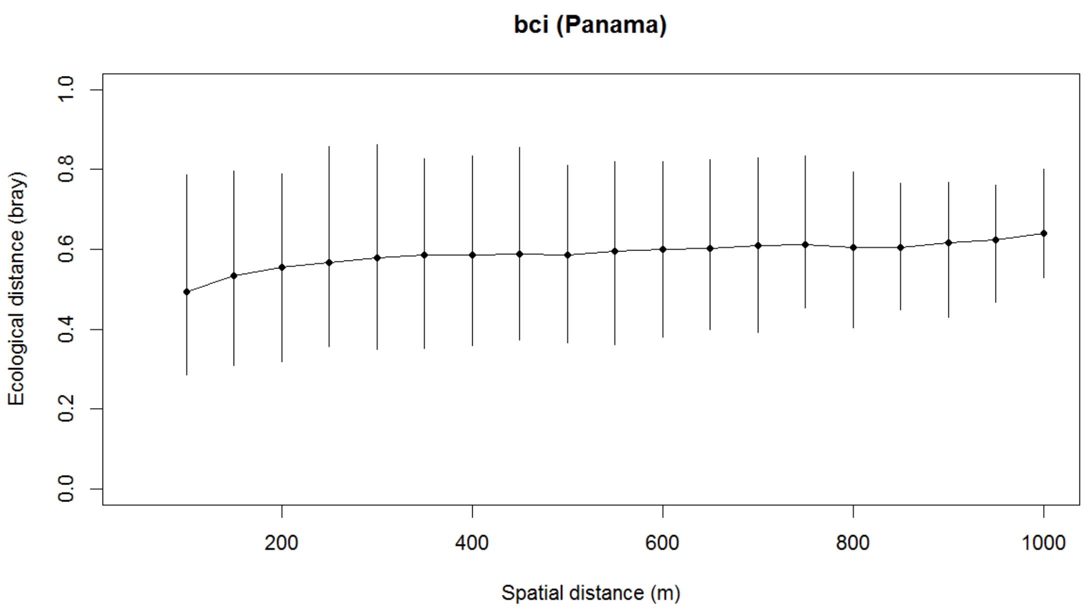

Figure SI3-3: Ecological distance vs. spatial distance in bci, Panama

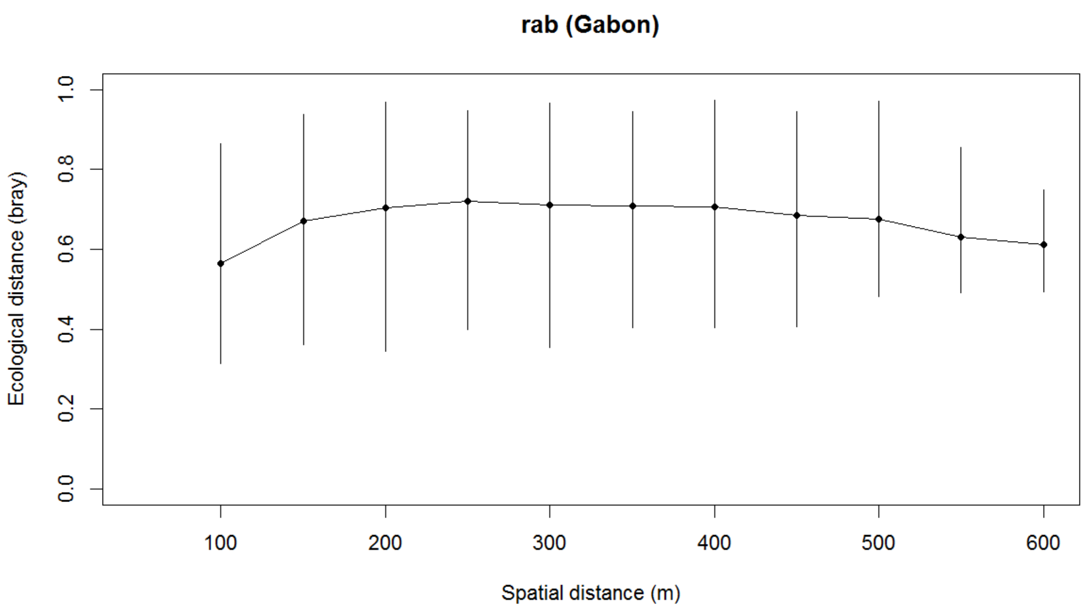

Figure SI3-4: Ecological distance vs. spatial distance in rab, Gabon

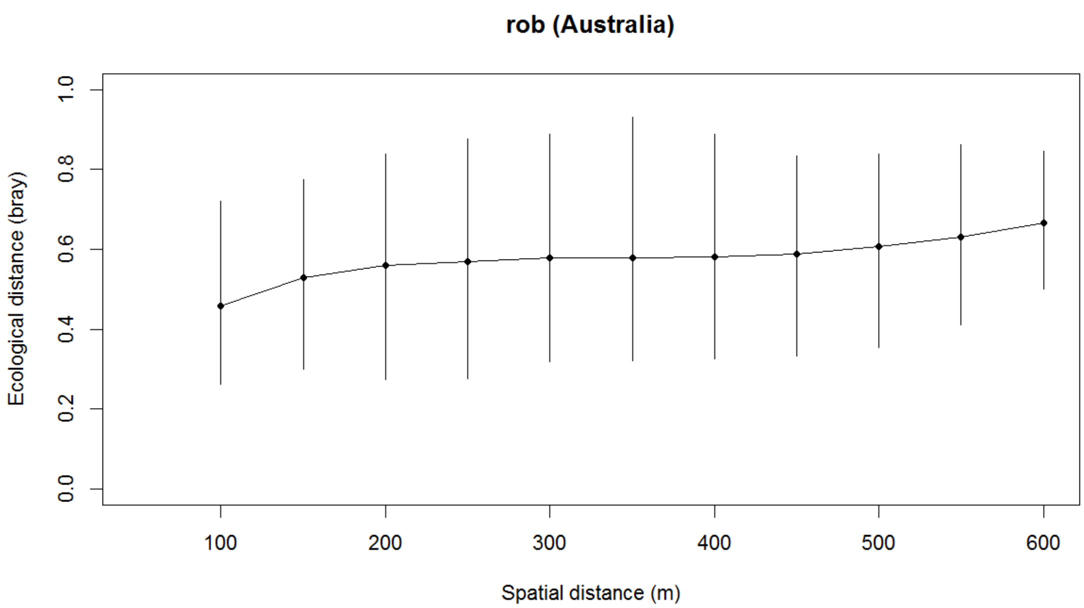

Figure SI3-5: Ecological distance vs. spatial distance in rob, Australia

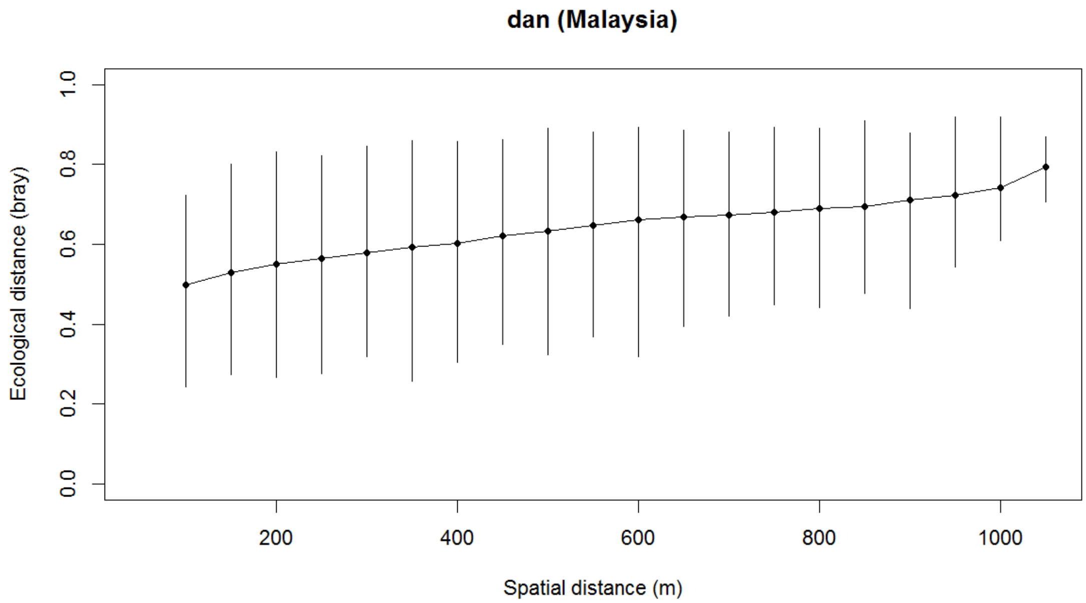

Figure SI3-6: Ecological distance vs. spatial distance in dan, Malaysia

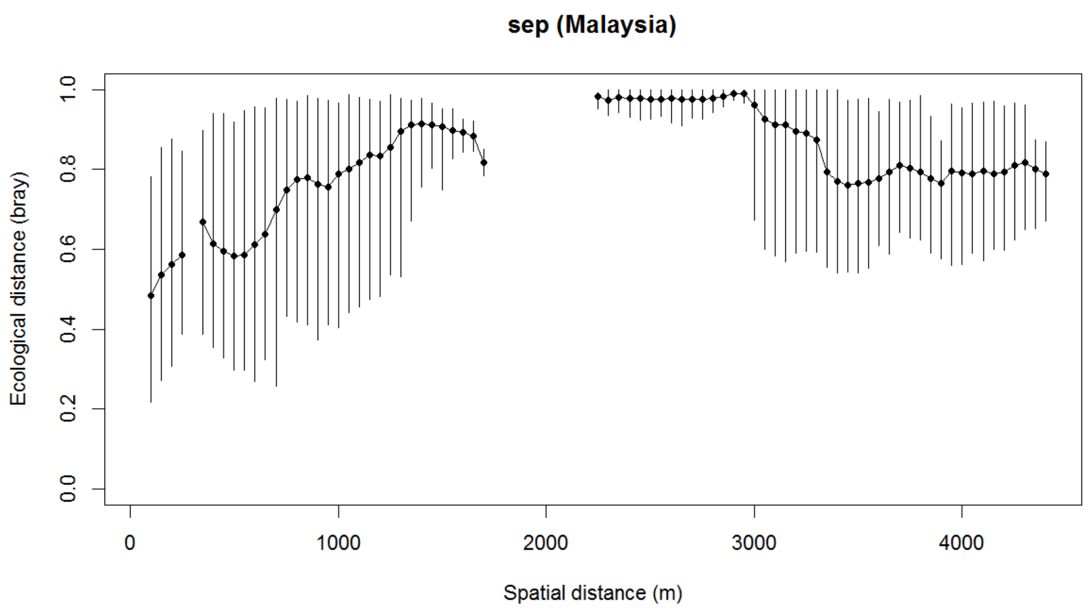

Figure SI3-7: Ecological distance vs. spatial distance in sep, Borneo

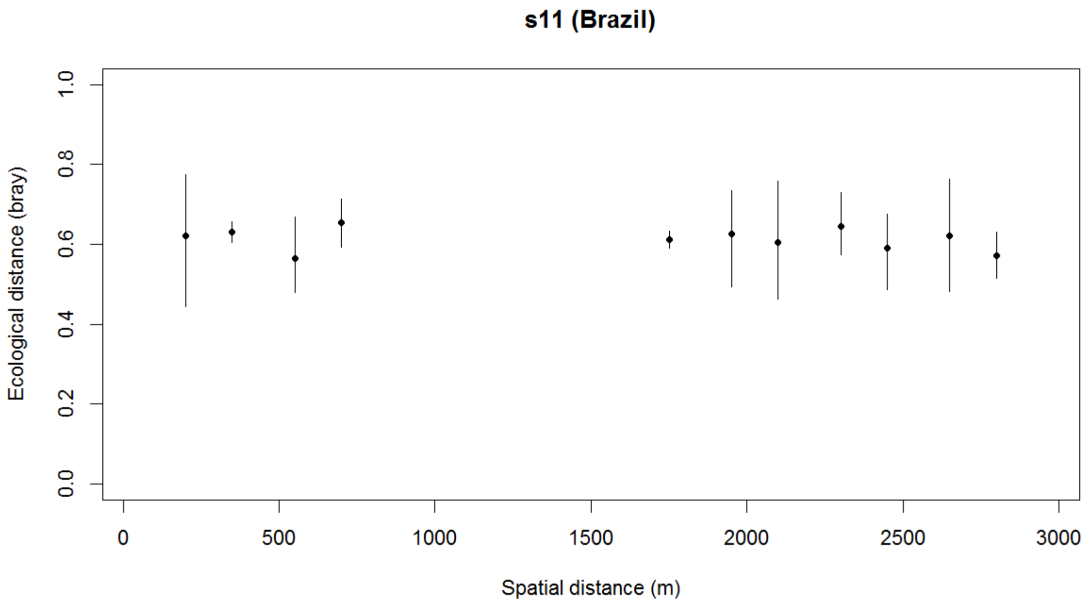

Figure SI3-8: Ecological distance vs. spatial distance in s11, Brazil

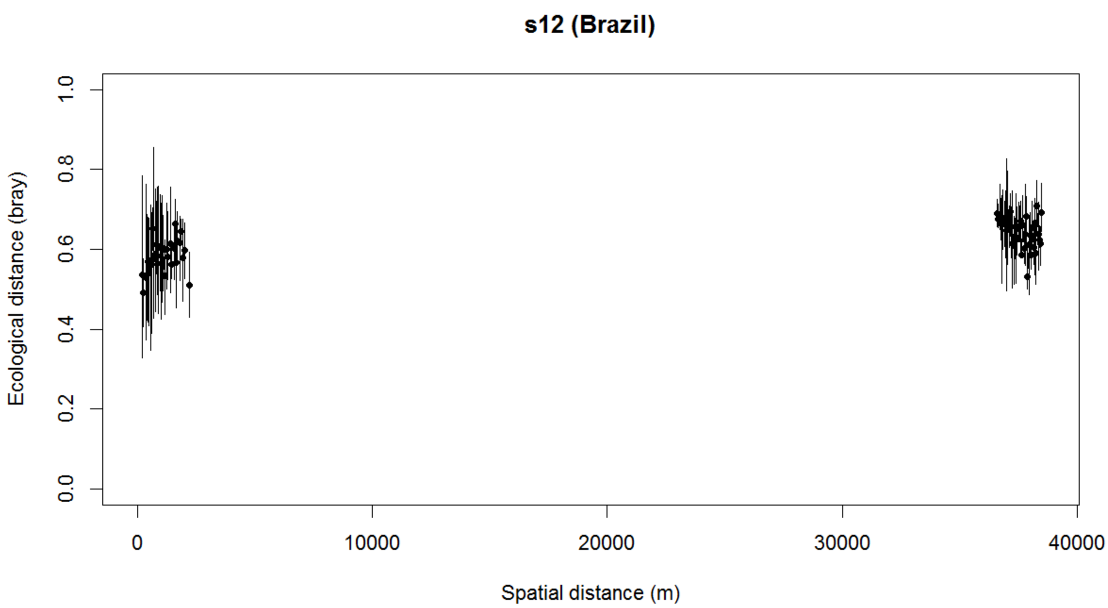

Figure SI3-9: Ecological distance vs. spatial distance in s12, Brazil

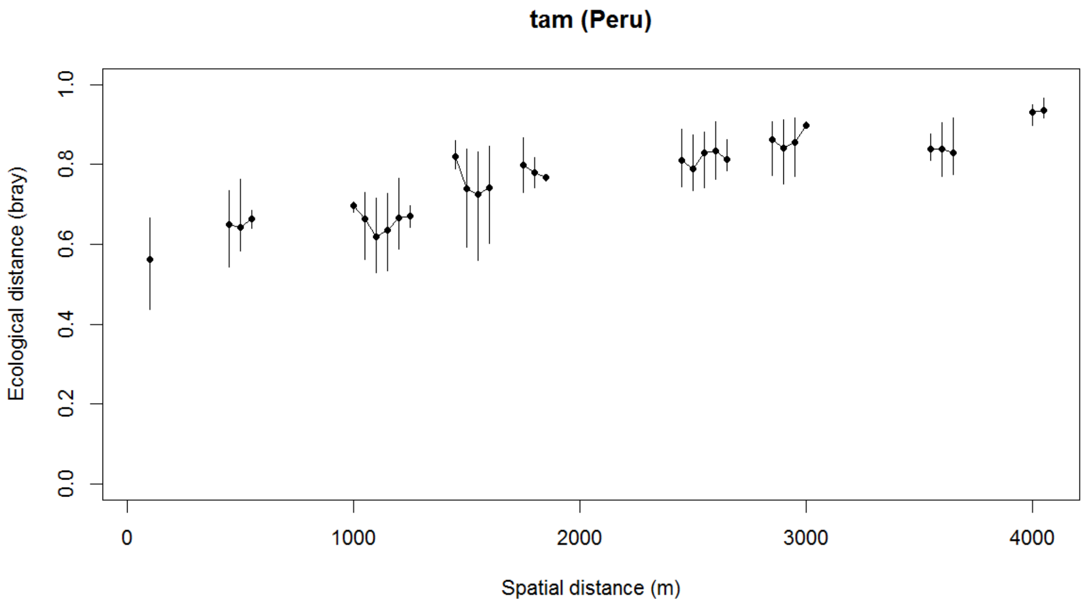

Figure SI3-10: Ecological distance vs. spatial distance in tam, Peru

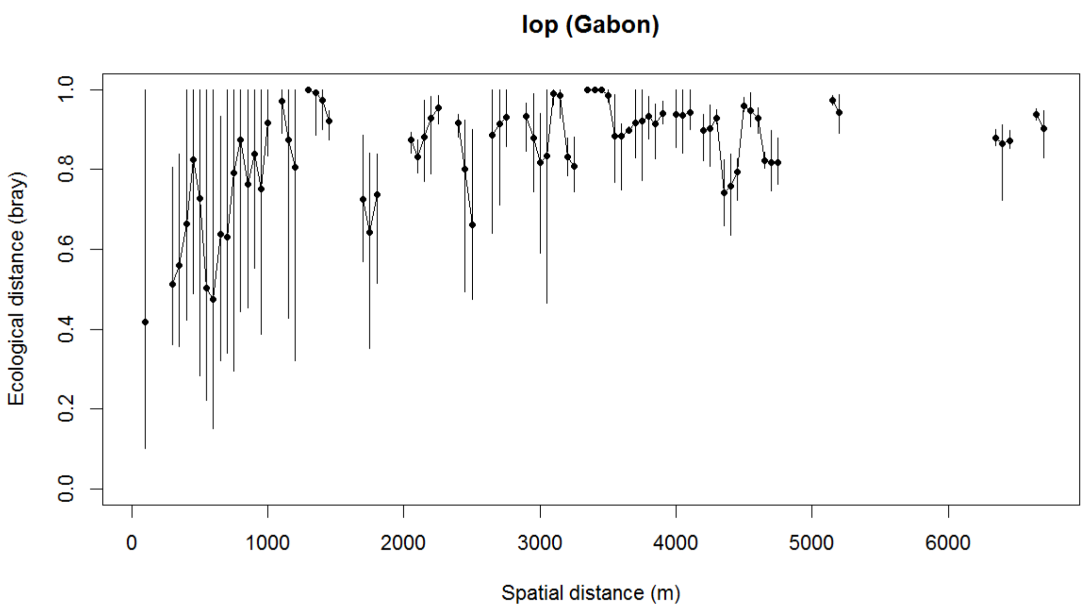

Figure SI3-11: Ecological distance vs. spatial distance in lop, Gabon

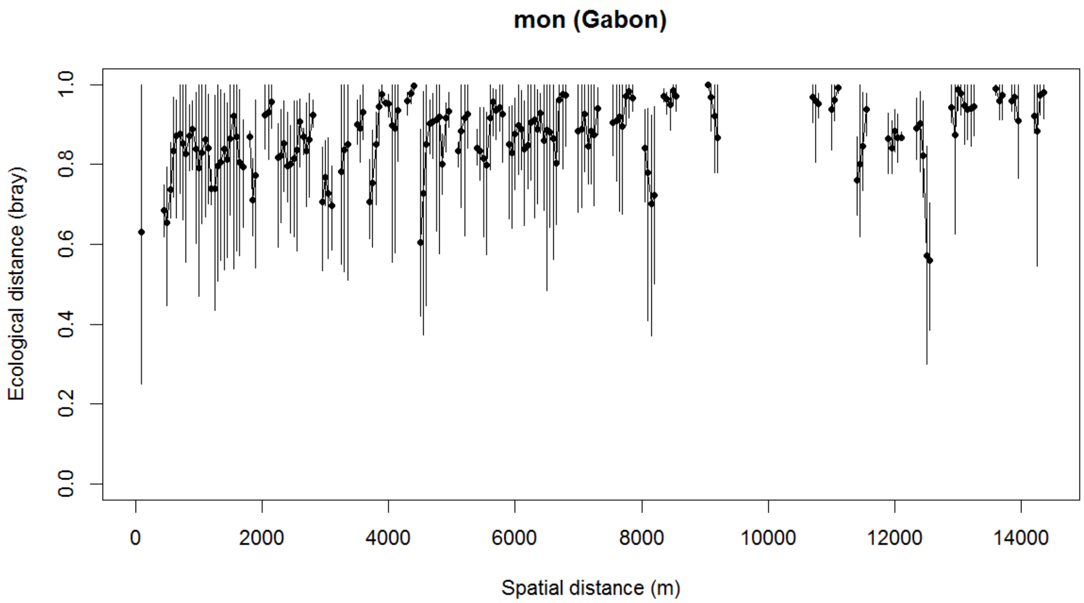

Figure SI3-12: Ecological distance vs. spatial distance in mon, Gabon

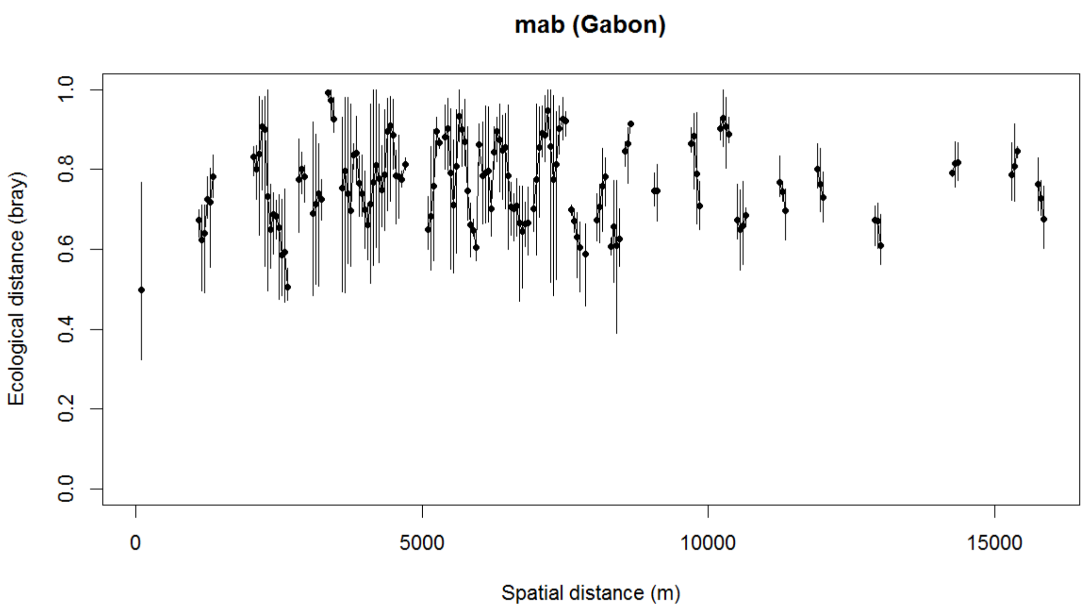

Figure SI3-13: Ecological distance vs. spatial distance in mab, Gabon

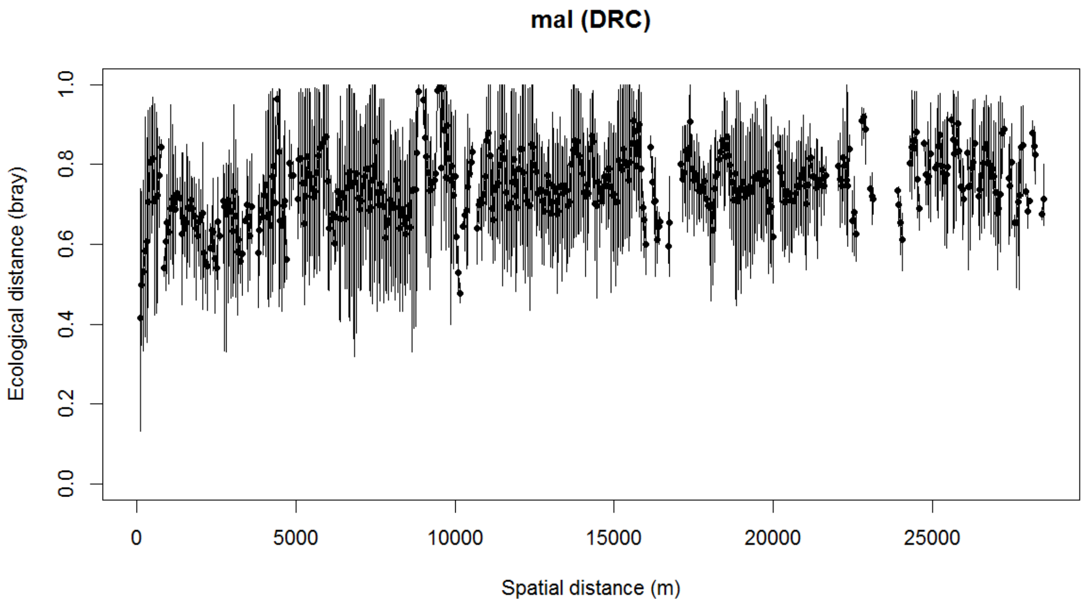

Figure SI3-14: Ecological distance vs. spatial distance in mal, DRC

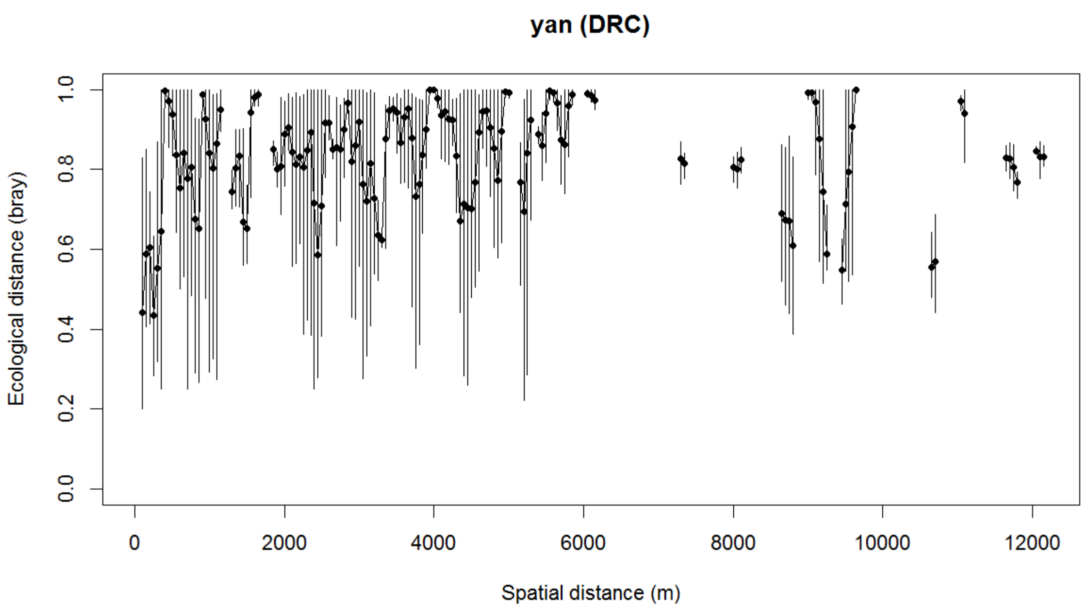

Figure SI3-15: Ecological distance vs. spatial distance in yan, DRC

SI4

Examples of observed vs. predicted plots from the global analysis at three spatial resolutions. The plots illustrate that the *tam* plots (orange dots) are a major outlier in the global predictions at the 1 ha resolution. At the 0.0625 ha resolution this study site is no longer an outlier as observed values of species richness are less extreme and within the range of values of the other datasets.

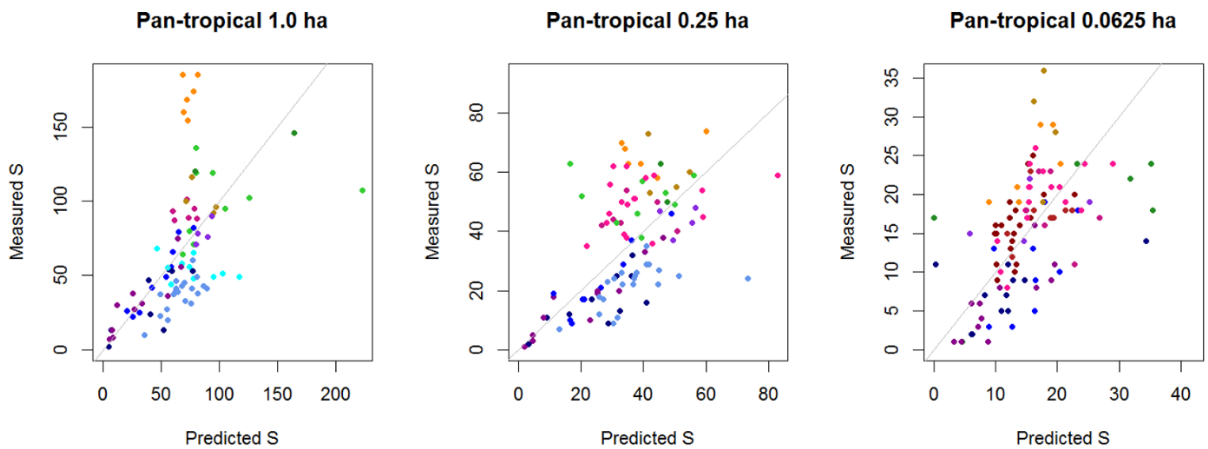

Figure SI4-1: Examples of predicted vs. measured tree species richness/area (S) using cross-validated global models from one randomly selected output of a Monte-Carlo simulation at each of the three spatial resolutions. Colors of points coincide with colors used in Figure 1 of the paper.