Supporting information for

"Current forest-savanna transition in northern South America departs from typical climatic thresholds"

Santiago Valencia[1,2*], Juan F. Salazar[1], Natalia Hoyos[3], Dolors Armenteras[4], Juan Camilo Villegas[2]

[1] Grupo GIGA, Escuela Ambiental, Facultad de Ingeniería, Universidad de Antioquía, Medellín, Colombia.
[2] Grupo de Investigación en Ecología Aplicada, Escuela Ambiental, Facultad de Ingeniería, Universidad de Antioquía, Medellín, Colombia
[3] Department of History and Social Sciences, Universidad del Norte, Barranquilla, Colombia
[4] Grupo de Ecología del Paisaje y Modelación de Ecosistemas ECOLMOD, Departamento de Biología, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Contents of this file:

Supporting information S1

The Global Ecosystem Dynamics Investigation (GEDI) is a new spaceborne LiDAR instrument aboard the International Space Station (ISS) that collects vegetation structure data at an average footprint resolution of 25 m since April 2019 for a nominal two-years mission (Dubayah and others 2020). GEDI consists of 3 lasers that reach 25 m diameter on the ground (footprints), resulting in 8 tracks of data, spaced at 60 m along track and 600 m across track within a ~4.2 km swath (Supporting Information Figure S1a). Gridded products based solely on GEDI sample data will have a spatial resolution of at least 1 km. However, those products are still in development (Luthcke and others 2021). GEDI L2B product data was collected between April 19, 2019 and September 02, 2020. In GEDI, Lidar waveforms quantify the vertical distribution of vegetation by recording the amount of laser energy reflected by plant material including stems, branches, and leaves at different heights above the ground at a footprint resolution of 25 m (Dubayah and others 2020). For each footprint, we extracted the percentage canopy cover and Plant Area Volume Density (PAVD) profile using the rGEDI R package (Silva and others 2020). Canopy cover represents the percent of the ground covered by the vertical projection of canopy material (i.e., leaves, branches and stems only) considering the gaps between and within the canopy (Tang and others 2019). The Plant Area Volume Density (PAVD) incorporates branches and trunk throughout the vertical profile of vegetation, in addition to leaves. For each footprint PAVD profile, we calculated the maximum PAVD ($PAVD_{max}$) as a proxy of vertical vegetation structure (Calders and others 2014; Decuyper and others 2018; Meeussen and others 2020). We excluded low-quality footprints (i.e., poor geolocation performance, waveforms of poor signal quality, and waveforms affected by cloud and other land surface conditions) according to the quality flags of GEDI L2B (Tang and others 2019). A detailed description of the GEDI products can be found in Tang and others (2019) and Dubayah and others (2020).

To obtain gridded GEDI data for the study area, we averaged the footprints within each 0.01° x 0.01° pixel as suggested by Rishmawi and others (2021). Given that the number of GEDI observations within each pixel ranged between 1 and 60 (see Supporting Information Figure S1b), we excluded pixels calculated with less than 20 footprints to minimize the effect of under-sampling observations within pixels while maintaining a large number of available pixels (Figure S1c in Supporting Information). Additionally, the results do not show substantial changes in mean ($\overline{X}$) or variance (S^2) when considering different footprint thresholds; for example, the maximum and S^2 PAVD values when considering a threshold of 10 were ($\overline{X}_{PAVDmax}$ = 0.362 and $S^2_{PAVDmax}$ = 0.057), compared to a threshold of 30 ($\overline{X}_{PAVDmax}$ = 0.384 and $S^2_{PAVDmax}$ = 0.059) or a threshold of 20 ($\overline{X}_{PAVDmax}$ = 0.371 and $S^2_{PAVDmax}$ = 0.058). The percentage tree cover from MODIS has been used in a wide variety of ecological research including vegetation modeling, estimating restoration potential, and identifying forest-savanna bimodality (Adzhar and others 2021). Percent tree cover describes the percent of each pixel covered by canopy cover of trees above 5 m in height (DiMiceli *et al*., 2015). In this case, we used the percent tree cover available for the year 2019 based on MODIS inputs dating from 6 March 2019 to 6 March 2020 at a 250 m resolution.

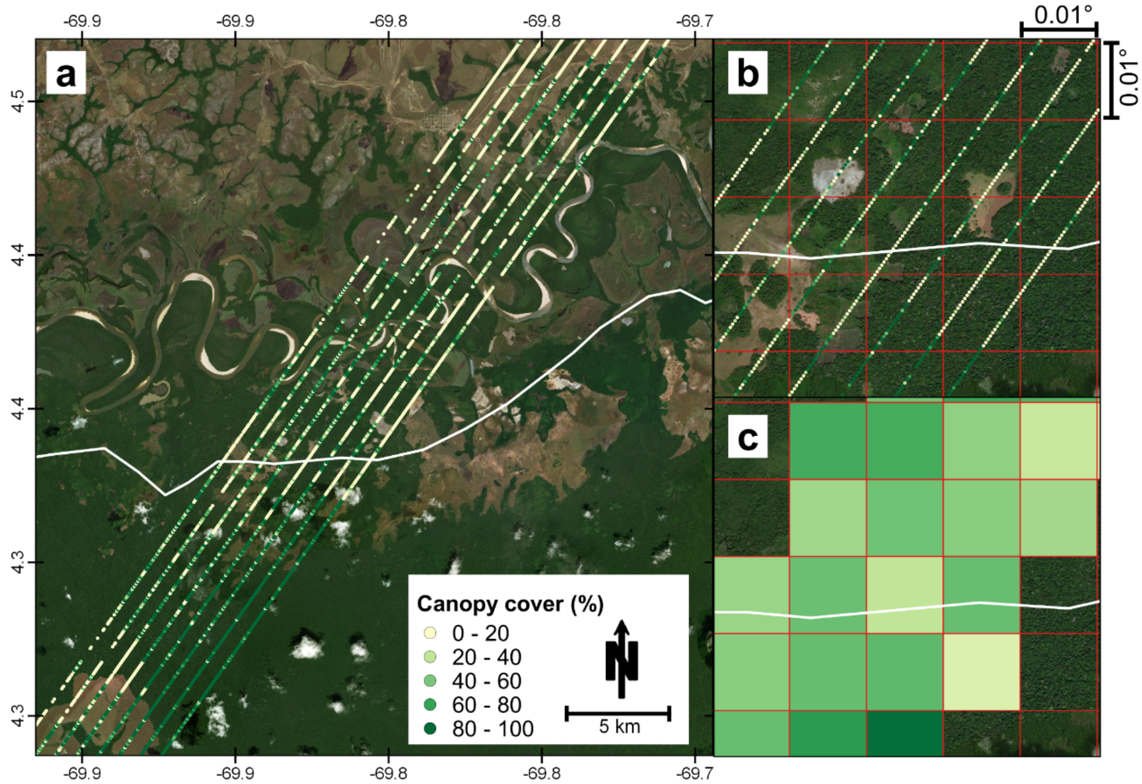

Figure S1. a) Sample GEDI beam ground transects along forest-savanna transition and b) the number of GEDI footprints available in a pixel at 0.01° x 0.01° resolution (red lines). c) Result of average GEDI footprints in each pixel, excluding pixels with less than 20 footprints. The white line represents the boundary of the Llanos ecoregion. Background image from Google Earth.

Pixels were defined as forest or savanna based on each vegetation structure descriptor: canopy cover, tree cover, and $PAVD_{max}$. We used a threshold of 60% tree cover to define savanna (< 60%) and forest (≥60%) as suggested by Hirota and others (2011). Given the difference between tree cover (DiMiceli and others 2015) and canopy cover (Tang and others 2019) as well as between MODIS and GEDI (e.g., MODIS tree cover saturates at ~80%, while the GEDI canopy cover data maxes out at 100%), we used a threshold of 40% canopy cover to classify pixels as savanna (< 40%) or forest (≥ 40%). For $PAVD_{max}$, we defined forest and savanna using a threshold of 0.10 m^2/m^3 following the variable distribution and previous values reported for tropical forests (e.g., Decuyper and others 2018). Finally, the selected thresholds also ensured a similar number of pixels classified as forest and savanna using the different vegetation structure descriptor variables (as indicated in Figure S2). However, further studies are required due to GEDI LiDAR limitations to characterize short stature (< 3 m) and discontinuous vegetation such as those found in savanna areas (Li and others 2023).

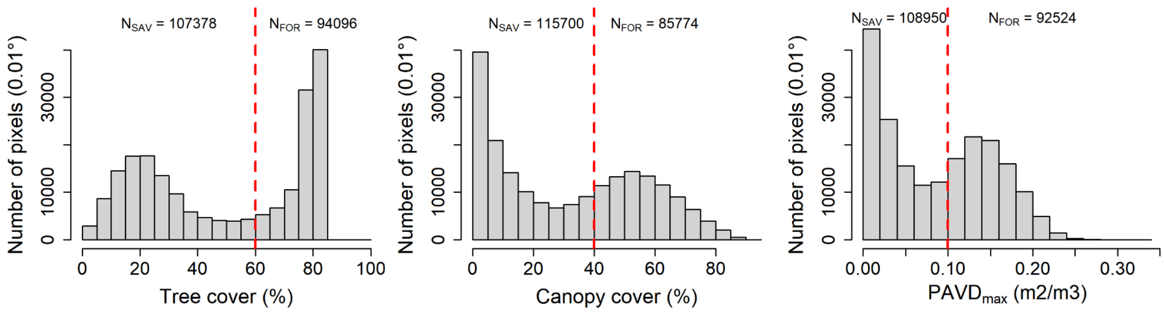

Figure S2. Tree cover, canopy cover, and $PAVD_{max}$ distribution. The vertical red line indicates the threshold used to define forest (N_{FOR}) and savanna (N_{SAV}) pixels for each vegetation structure descriptor variable.

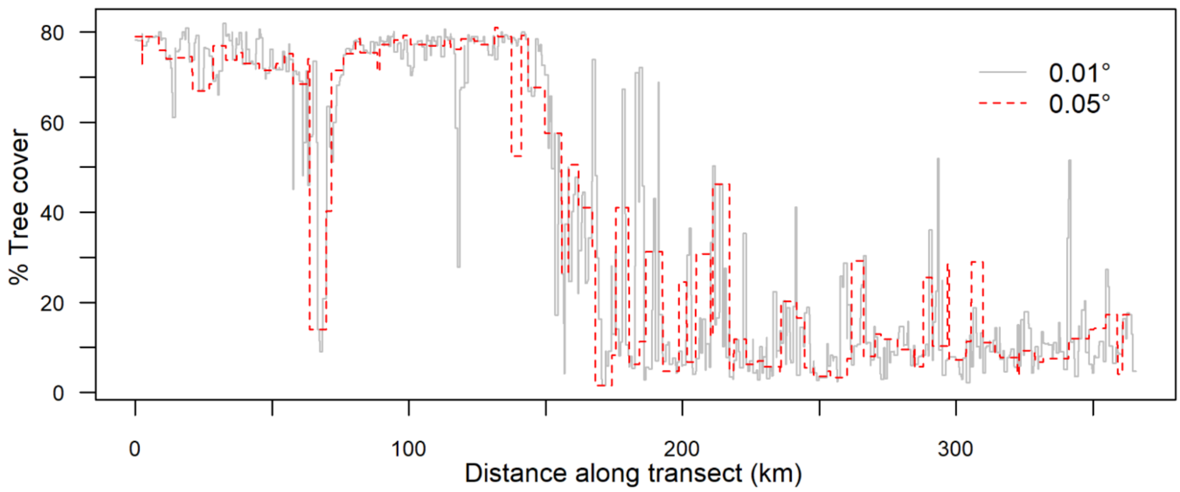

Figure S3. Example of differences in percentage tree cover from MODIS along a forest-savanna transect at 0.01° (solid gray line) and 0.05° (dotted red line) spatial resolution.

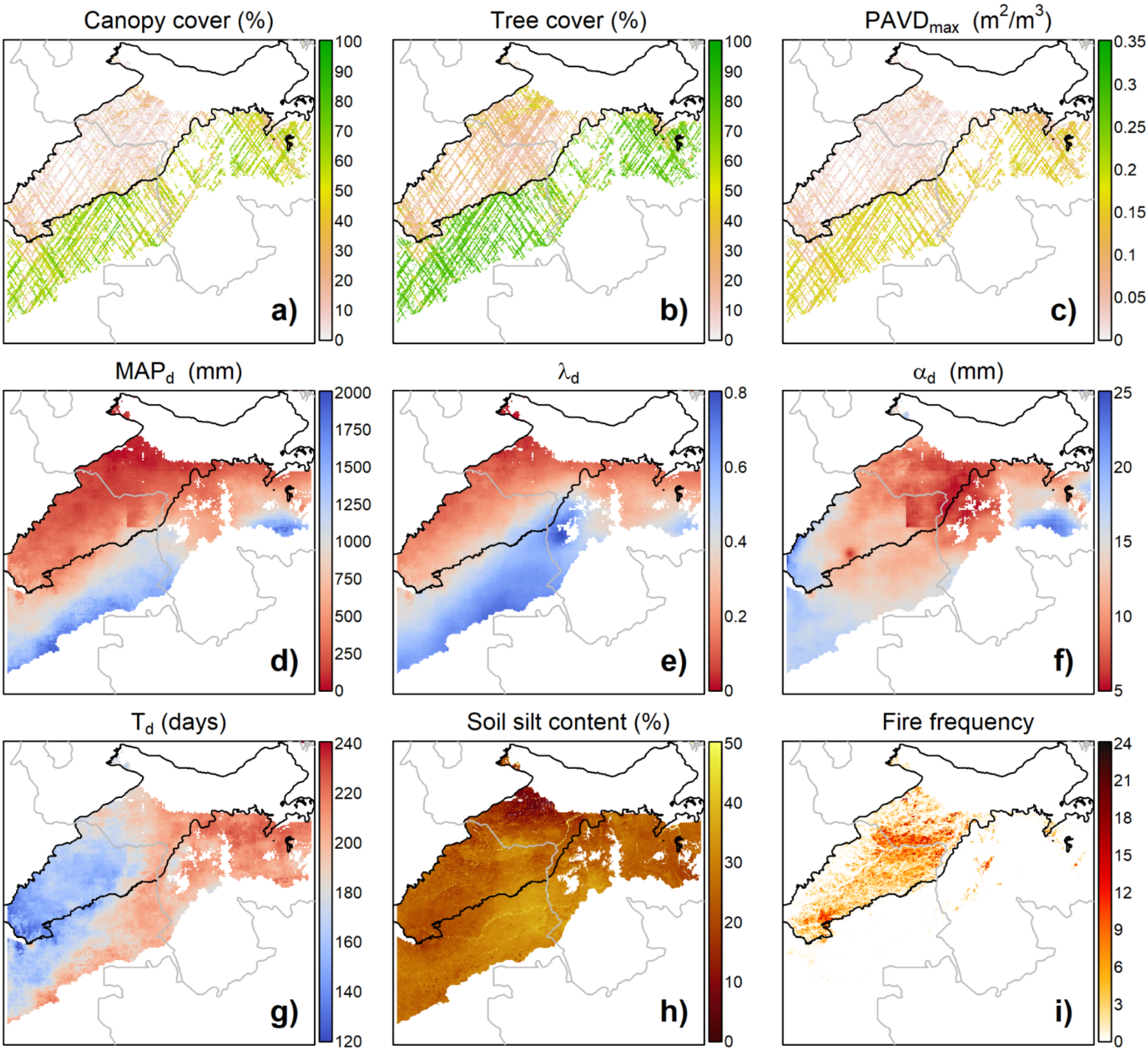

Figure S4. Spatial maps of predictor variables. The black outlined polygon represents the Llanos Ecoregion. Our study area is defined by the colored pixels. Pixels outside the study area were excluded following the criteria described in Study Area Delimitation and Forest-Savanna Discrimination Section.

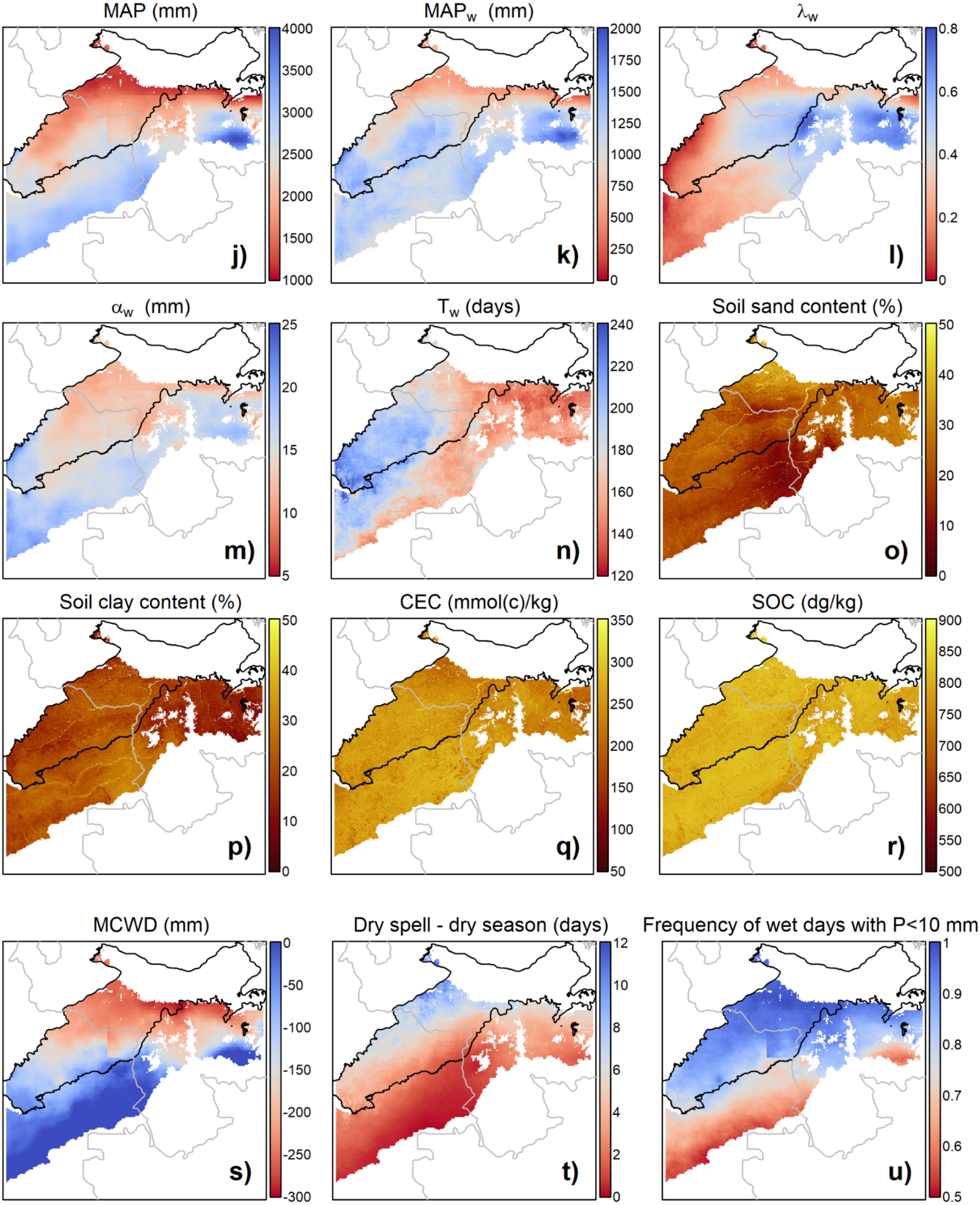

Figure S4. Spatial maps of predictor variables. (Continued)

Supporting information S2

Given that the length of transects varied depending on the limits of the study area (Figure 1) and GEDI available pixels (Figure S3a in Supporting Information Figure), we excluded median and percentiles values calculated with less than 375 pixels (i.e., 50% of GEDI available pixels along all transects analyzed).

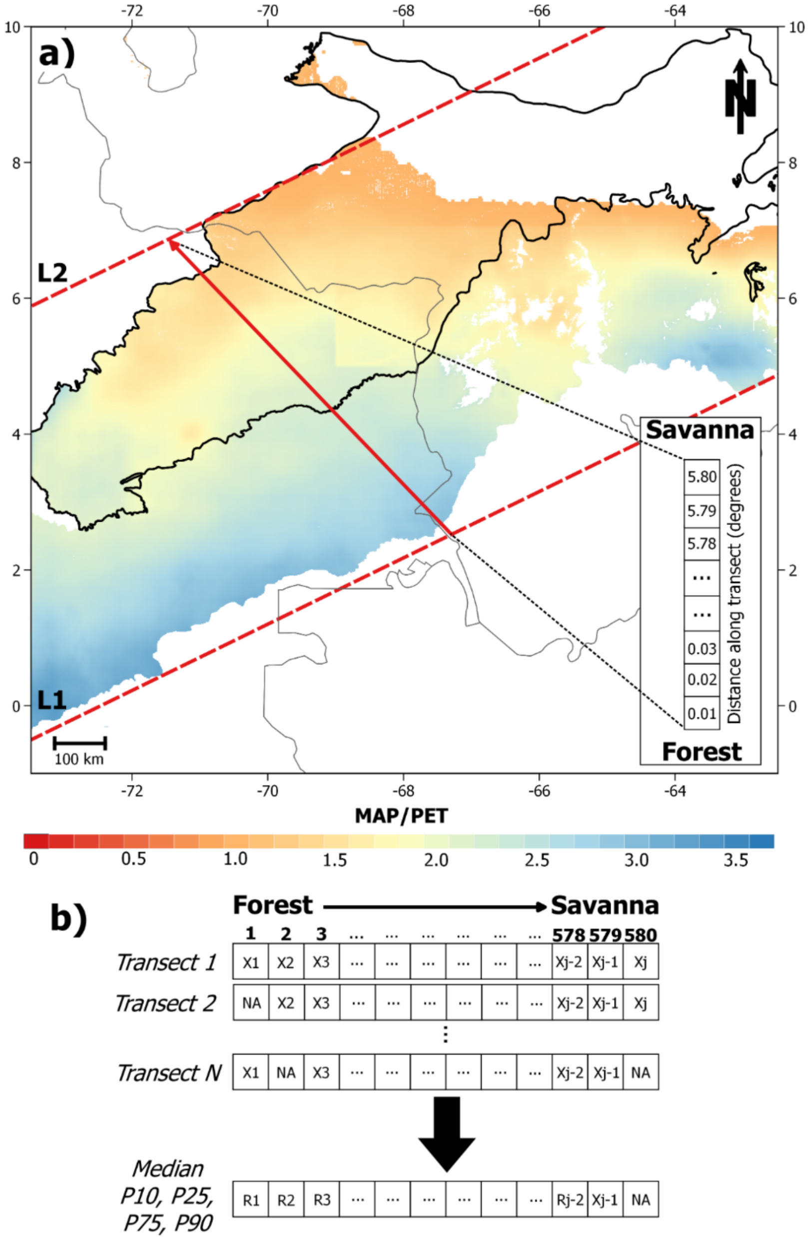

Figure S5. a) Diagram of transect analyzed along the forest-savanna transition. Forest and savanna limits were defined based on limits of ecoregions. Transects were 0.01° (~ 1 km) wide, of variable length, and were sampled in 0.01° increments (0.01° x 0.01° pixels). b) Transects were aligned based on the forest-savanna edge, and summary statistics were subsequently calculated for each pixel.

Supporting information S3

We estimated the importance of each predictor variable to explain response variables based on the Pearson correlation coefficient (r) and the Spearman rank correlation analysis (r_s), indicating whether relations are linear or nonlinear, respectively. Only those variables with correlations higher than |0.2| and p-value < 0.01 were considered for further analysis. Potential collinearity between predictor variables was also assessed with those metrics (r or r_s ≥ |0.70|; (Dormann and others 2013).

As expected, canopy cover, tree cover, and $PAVD_{max}$ were highly correlated with each other (r_s and r ≥ 0.82), particularly canopy cover and $PAVD_{max}$ that come from the same dataset. However, we observed differences in the representation of vegetation structure between these variables (Supporting Information Figure S6). For example, although canopy cover and $PAVD_{max}$ show almost a perfect fit for the linear regression (the regression $R^2 = 0.97$), the dispersion of points was lower for low canopy cover (< 25%) and $PAVD_{max}$ (< 0.1 m^2/m^3) than for high values of this vegetation descriptors (canopy cover > 50%; $PAVD_{max}$ > 0.2 m^2/m^3). This contrast evidenced the differences between horizontal (i.e., canopy cover) and vertical (i.e., $PAVD_{max}$) vegetation structure as shown in Supporting Information Figure S7a. Additionally, as reported by DiMiceli and others (2021), GEDI's values are, on average, higher and lower than tree cover from MODIS at the higher and lower end of the range, respectively (Supporting Information Figure S6b), which is related to differences in these products' spatial resolution and methods. Therefore, our analysis included canopy cover, tree cover, and $PAVD_{max}$ and considers that the most robust patterns were those consistent across the three vegetation descriptors.

Our results suggested that precipitation during the dry season is a better predictor of forest-savanna transition than precipitation during the wet season. MAP_d (r_s and r ≥ 0.64) and λ_d (r_s and r ≥ 0.54), as well as other PV components during the dry season (r_s and r ≥ 0.26; Supporting Information Tables S1 and S2), exhibit a high positive correlation with vegetation descriptors with both the Spearman (r_s) and Pearson (r) correlation metrics. For this reason, we focused our analysis on the dry season consistent with the previous results of Hoyos and others (2022) for the Llanos ecoregion.

Table S1. Median Spearman rank correlation (r_s) analysis for all predictor variables analyzed. “cover” = % canopy cover; “tree” = % tree cover; “PAVD” = maximum Plant Area Volume Density; “sand” = topsoil sand content; “clay” = topsoil clay content; “silt” = topsoil silt content; “SOC” = topsoil organic carbon; “CEC” = topsoil cation exchange capacity; “MAP” = mean annual precipitation; “$\boldsymbol{\lambda}$” = frequency of rainy days ($p > 0$); “$\boldsymbol{\alpha}$” = intensity of precipitation; “T” = length of wet or dry seasons; “fire” = fire frequency. “Ds_d” = the length of dry spells in the dry season; “$\boldsymbol{\lambda}_{10d}$” = frequency of wet days with precipitation < 10 mm/day within the dry season. d and w represent the dry and the wet season, respectively. Values are calculated using the selected 1000 random samples of 5% (N=10074) of data. The standard deviation of correlation values is lower than 0.05 for all variables.

	cover	tree	PAVD	sand	clay	silt	SOC	CEC	MAP	$\boldsymbol{\lambda}_w$	$\boldsymbol{\lambda}_d$	$\boldsymbol{\alpha}_w$	$\boldsymbol{\alpha}_d$	MAP_w	MAP_d	T_w	T_d	fire	Ds_d	$\boldsymbol{\lambda}_{10d}$
cover	1																			
tree	0.82	1																		
PAVD	0.99	0.82	1																	
sand	0.21	0.20	0.22	1																
clay	0.01	-0.01	0.01	-0.70	1															
silt	**-0.30**	**-0.29**	**-0.30**	-0.83	0.38	1														
SOC	0.01	0.01	0.01	-0.14	0.26	0.03	1													
CEC	-0.13	-0.11	-0.14	-0.59	0.49	0.51	0.33	1												
MAP	0.56	0.56	0.58	0.49	-0.24	-0.49	0.16	-0.33	1											
$\boldsymbol{\lambda}_w$	-0.01	-0.04	-0.04	0.02	0.09	-0.16	0.12	0.10	0.02	1										
$\boldsymbol{\lambda}_d$	**0.57**	**0.54**	**0.58**	0.64	-0.39	-0.60	0.08	-0.43	0.86	0.07	1									
$\boldsymbol{\alpha}_w$	0.53	0.52	0.54	0.30	-0.06	-0.34	0.19	-0.22	0.86	-0.24	0.65	1								
$\boldsymbol{\alpha}_d$	**0.42**	**0.44**	**0.43**	0.09	0.05	-0.11	0.18	-0.13	0.71	-0.44	0.38	0.81	1							
MAP_w	0.23	0.18	0.25	0.30	-0.09	-0.30	0.15	-0.26	0.74	0.13	0.47	0.64	0.58	1						
MAP_d	0.67	0.66	0.68	0.48	-0.21	-0.51	0.16	-0.30	0.94	0.01	0.92	0.82	0.64	0.56	1					
T_w	-0.26	-0.27	-0.26	0.15	-0.22	0.06	-0.14	-0.32	0.01	-0.52	-0.08	0.02	0.27	0.36	-0.18	1				
T_d	**0.26**	**0.27**	**0.26**	-0.15	0.22	-0.06	0.14	0.32	-0.01	0.52	0.08	-0.02	-0.27	-0.36	0.18	-1*	1			
fire	**-0.64**	**-0.64**	**-0.64**	-0.01	-0.07	0.15	-0.01	0.05	-0.38	0.11	-0.35	-0.42	-0.37	-0.08	-0.49	0.25	-0.25	1		
Ds_d	-0.61	-0.58	-0.62	-0.63	0.38	0.60	-0.09	0.41	-0.88	0.03	-0.99	-0.70	-0.45	-0.48	-0.95	0.10	-0.10	0.40	1	
$\boldsymbol{\lambda}_{10d}$	-0.62	-0.62	-0.63	-0.50	0.26	0.48	-0.12	0.37	-0.95	0.16	-0.88	-0.85	-0.76	-0.61	-0.96	-0.07	0.07	0.44	0.91	1

The bold values indicate the correlations between response variables and predictor variables selected for statistical model analysis. The shaded values indicate low (<|0.2|) and non-statistically significant ($p > 0.01$) correlation values. *T_w and T_d show a perfect negative correlation because the length of the wet or dry season is linearly dependent on each other (i.e., $T_w = 365 - T_d$).

Table S2. Median Pearson correlation (r) coefficient for all predictor variables analyzed. “cover” = % canopy cover; “tree” = % tree cover; “PAVD” = maximum Plant Area Volume Density; “sand” = topsoil sand content; “clay” = topsoil clay content; “silt” = topsoil silt content; “SOC” = topsoil organic carbon; “CEC” = topsoil cation exchange capacity; “MAP” = mean annual precipitation; “$\boldsymbol{\lambda}$” = frequency of rainy days ($p > 0$); “$\boldsymbol{\alpha}$” = intensity of precipitation; “T” = length of wet or dry seasons; “fire” = fire frequency. “Ds_d” = the length of dry spells in the dry season; “$\boldsymbol{\lambda}_{10d}$” = frequency of wet days with precipitation < 10 mm/day within the dry season. d and w represent the dry and the wet season, respectively. Values are calculated using the selected 1000 random samples of 5% (N=10074) of data. The standard deviation of correlation values is lower than 0.05 for all variables.

	cover	tree	PAVD	sand	clay	silt	SOC	CEC	MAP	$\boldsymbol{\lambda}_w$	$\boldsymbol{\lambda}_d$	$\boldsymbol{\alpha}_w$	$\boldsymbol{\alpha}_d$	MAP_w	MAP_d	T_w	T_d	fire	Ds_d	$\boldsymbol{\lambda}_{10d}$
cover	1																			
tree	0.87	1																		
PAVD	0.99	0.86	1																	
sand	0.24	0.22	0.25	1																
clay	0.03	0.04	0.02	-0.56	1															
silt	**-0.31**	**-0.28**	**-0.31**	-0.77	0.38	1														
SOC	0.02	0.03	0.02	-0.02	0.25	0.01	1													
CEC	-0.10	-0.06	-0.11	-0.51	0.53	0.54	0.35	1												
MAP	0.53	0.54	0.55	0.49	-0.18	-0.52	0.19	-0.32	1											
$\boldsymbol{\lambda}_w$	0.03	0.03	0.03	0.04	0.15	-0.16	0.16	0.11	0.10	1										
$\boldsymbol{\lambda}_d$	**0.57**	**0.58**	**0.58**	0.64	-0.38	-0.60	0.10	-0.40	0.82	0.04	1									
$\boldsymbol{\alpha}_w$	0.51	0.52	0.52	0.30	0.03	-0.36	0.16	-0.20	0.84	-0.25	0.62	1								
$\boldsymbol{\alpha}_d$	**0.39**	**0.42**	**0.40**	0.07	0.09	-0.15	0.15	-0.08	0.69	-0.37	0.34	0.80	1							
MAP_w	0.25	0.24	0.27	0.37	-0.08	-0.41	0.15	-0.30	0.84	0.21	0.49	0.66	0.58	1						
MAP_d	0.64	0.67	0.65	0.47	-0.21	-0.48	0.17	-0.25	0.89	-0.01	0.90	0.78	0.62	0.50	1					
T_w	-0.27	-0.31	-0.26	0.09	-0.22	0.03	-0.14	-0.30	0.06	-0.53	-0.10	0.05	0.28	0.35	-0.20	1				
T_d	**0.27**	**0.31**	**0.26**	-0.09	0.22	0.03	0.14	0.30	-0.06	0.53	0.10	-0.05	-0.28	-0.35	0.20	-1*	1			
fire	**-0.51**	**-0.52**	**-0.51**	0.02	-0.08	0.06	-0.03	0.03	-0.26	0.12	-0.28	-0.32	-0.33	-0.04	-0.37	0.17	-0.17	1		
Ds_d	-0.60	-0.58	-0.61	-0.63	0.30	0.63	-0.13	0.39	-0.87	-0.08	-0.96	-0.70	-0.42	-0.61	-0.88	0.07	-0.07	0.29	1	
$\boldsymbol{\lambda}_{10d}$	-0.61	-0.64	-0.62	-0.47	0.25	0.46	-0.12	0.31	-0.90	0.18	-0.87	-0.82	-0.72	-0.57	-0.96	-0.03	0.03	0.36	0.86	1

The bold values indicate the correlations between response variables and predictor variables selected for statistical model analysis. The shaded values indicate low (<|0.2|) and non-statistically significant ($p > 0.01$) correlation values. *T_w and T_d show a perfect negative correlation because the length of the wet or dry season is linearly dependent on each other (i.e., $T_w = 365 - T_d$).

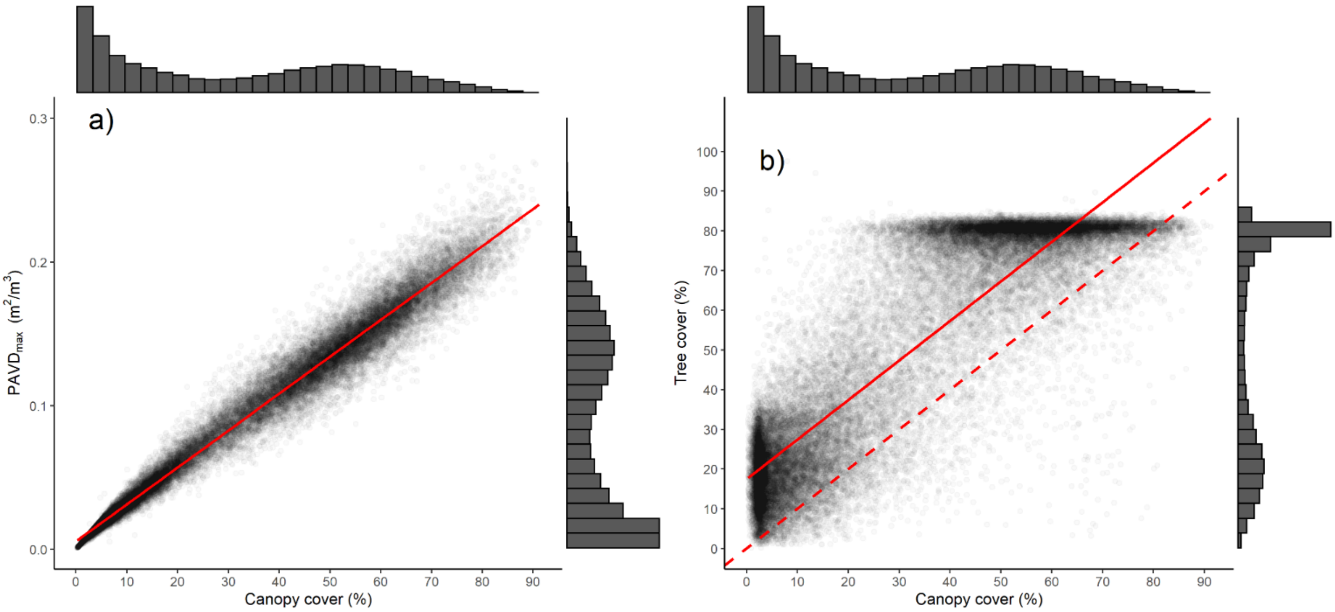

Figure S6. Comparison of a) maximum PAVD (PAVD$_{max}$= 0.559 + 0.257*canopy cover; the regression R^2 = 0.97), and b) tree cover (tree cover = 17.463 + 0.996*canopy cover; the regression R^2 = 0.76, Root mean squared difference = 22.10) with canopy cover. The solid and dashed red line represents the linear regression line fit and the 1:1 line, respectively. Black points show a subset (20%, N = ~40000) of observations for purposes of visual clarity.

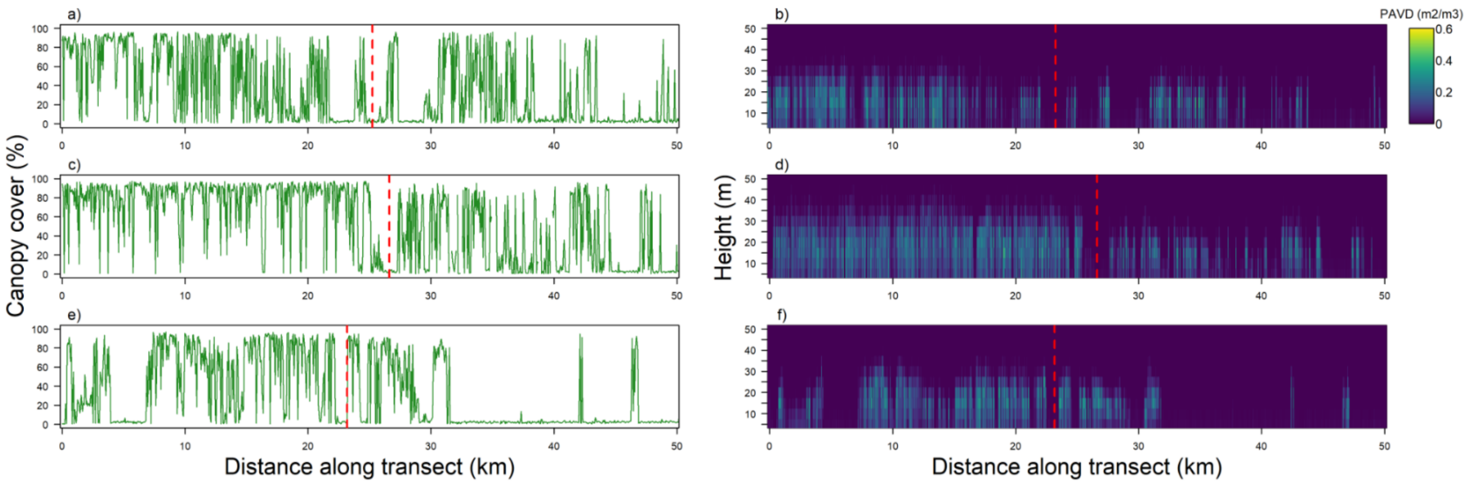

Figure S7. Example of a) % canopy cover and b) PAVD profile variations across three forest-savanna transects. The vertical red lines represent the transition between forest and savanna according to the limits of the Llanos ecoregion for each transect.

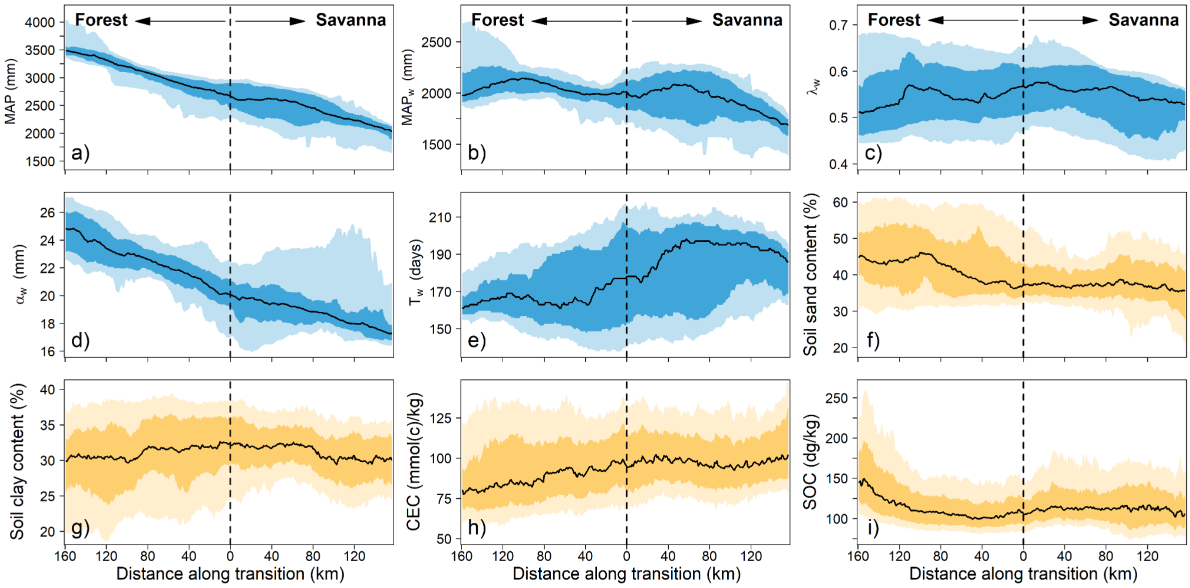

Figure S8. Changes in (a-e) climate and (f-i) variables across the forest-savanna transition. a) mean annual precipitation (MAP, mm), b) mean total wet-season precipitation (MAP_w), c) frequency of wet days (precipitation > 0) within the wet season (λ_w), d) intensity of wet days within the wet season (α_w), e) length of the wet season (T_w), f) topsoil (0-30 cm) sand content, g) topsoil (0-30 cm) clay content, f) topsoil (0-30 cm) cation exchange capacity (CEC), and i) topsoil (0-30 cm) soil organic carbon (SOC). The boundary between forest and savanna (vertical dashed line) along the transect is located where the mean of canopy cover, tree cover, or $PAVD_{max}$ (solid black line) crosses the forest - savanna threshold. In all panels, the lightest shaded band represents the 10^{th} - 90^{th} percentile, the darker shaded band is the 25^{th} - 75^{th} percentile, and the black line represents the median (50^{th} percentile) for all transects (n=1835).

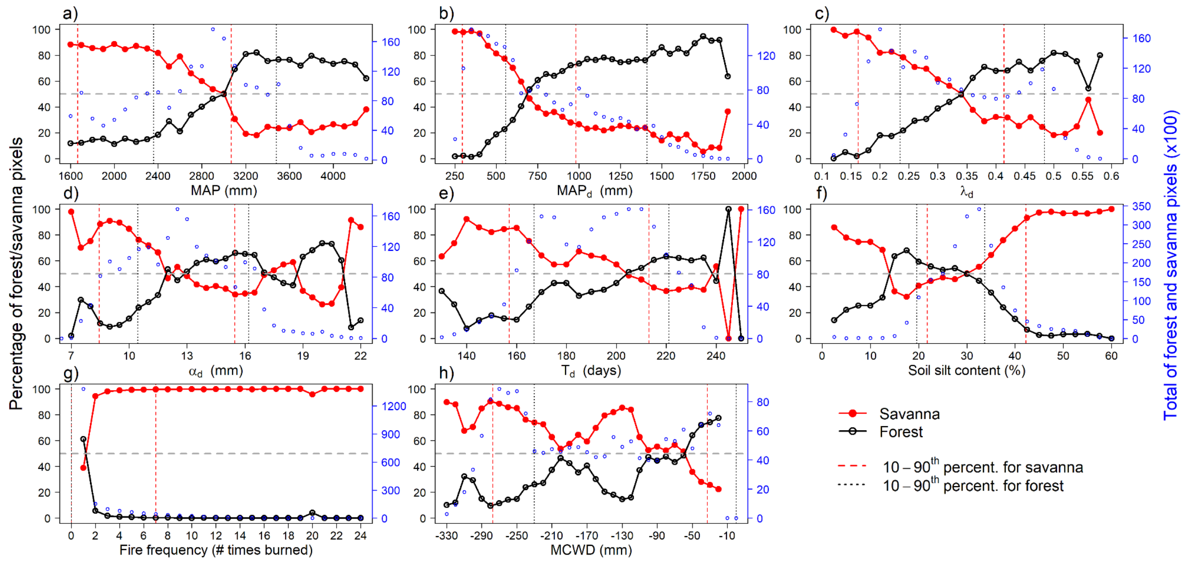

Figure S9. Percentage of forest and savanna pixels for different ranges of a) mean annual precipitation (MAP), b) mean total dry-season precipitation (MAP_d), c) frequency of wet days (precipitation > 0) within the dry season (λ_d), d) intensity of wet days within the dry season (α_d), e) length of the dry season (T_d), f) soil silt content, g) fire frequency, and h) maximum climatological water deficit (MCWD). Pixels were classified as forest and savanna based on canopy cover percentage (< 40% savanna; ≥ 40% forest). Vertical lines indicate each variable's 10^{th} and 90^{th} percentiles for savanna (red) and forest (black).

Supporting information S4

Following D'Onofrio and others (2018, 2019), we assumed binomial error distributions with a logit in GLMs, because vegetation descriptors are between 0 and 1 (i.e., canopy cover and tree cover were considered as fractions). In the case of $PAVD_{max}$, which potentially have values greater than 1, was remapped to the range [0:1] before using a logit function $(PAVD_{max})' = \frac{PAVD_{max} - min(PAVD_{max})}{max(PAVD_{max}) - min(PAVD_{max})}$. Due to PV components analyzed are not independent of each other and potential collinearity issues among the other predictor variables, we removed the selected predictor variable(s) with correlation (r or r_s) higher than ± 0.70 (Dormann and others 2013). In this sense, MAP_d and λ_d were the only selected predictor variables highly correlated (r and $r_s > 0.90$; Supporting Information Tables S1 and S2). For this reason, we left out MAP_d to maintain variable independence in GLMs. Although MAP_d was higher correlated with vegetation descriptors than λ_d (Supporting Information Tables S1 and S2), we selected λ_d instead MAP_d because precipitation frequency provided more information about precipitation variability during the dry season than total precipitation. Additionally, considering λ_d, α_d, and T_d, we indirectly considered the effect of MAP_d. We also did not include the length of dry spells (Ds_d) and frequency of wet days with precipitation (λ_{10d}) due to these variables were highly correlated with λ_d and α_d ($> \pm 0.72$) as well as each other (> 0.86; Supporting Information Tables S1 and S2).

Previous studies have also used specific samples of the data for statistical analysis when input data had different spatial resolutions. In this case, we used stratified--by deciles of canopy cover, tree cover, or $PAVD_{max}$-- random samples of 5% instead of 0.1% (Hirota and others 2011) or 1% (Staal and others 2016; Bernardi and others 2019) of all pixels. For each of 1000 samples by response variable and corresponding GLMs, we extracted the estimate (β) and standard error (σ) of each predictor variable and calculated a normal distribution with mean β and standard deviation σ as suggested by Compagnoni and others (2021). We used the resulting 1000 normal distributions of each predictor variable to estimate its median effect (i.e., direction and strength) and the 95% confidence interval (i.e., variability and significance) on each response variable.

Table S3. Generalized linear models (GLMs) estimates for standardized parameters and 95% confidence interval. The values of deviance explained represent the lowest and highest values from corresponding 1000 GLMs by each vegetation descriptor variable. Fire frequency (fire), mean daily precipitation frequency in the dry season (λ_d), mean daily precipitation intensity in the dry season (α_d), mean length of dry season (T_d), soil silt content (silt). * Predictor variables that have non-statistically significant effects.

Response variable	Predictors	Estimate	95 % Confidence interval	% Deviance explained (min and max)	% Deviance explained by each variable (min and max)
Canopy cover	fire	-0.79	[-1.00, -0.59]	52.21 - 55.40	32.61 - 35.83
	λ_d	0.40	[0.28, 0.53]		27.53 - 31.46
	α_d	0.25	[0.14, 0.35]		12.24 - 15.58
	T_d	0.26	[0.16, 0.36]		5.56 - 8.55
	silt	-0.04*	[-0.16, 0.08]		7.33 - 10.28
Tree cover	fire	-0.42	[-0.54, -0.30]	54.15 - 57.97	26.21 - 30.57
	λ_d	0.48	[0.36, 0.60]		29.74 - 33.74
	α_d	0.37	[0.26, 0.48]		14.53 - 18. 99
	T_d	0.37	[0.28, 0.47]		6.79 - 11.03
	silt	0.04*	[-0.07, 0.15]		5.46 - 9.52
$PAVD_{max}$	fire	-0.68	[-0.89, -0.47]	52.75 - 56.39	32.14 - 37.17
	λ_d	0.36	[0.23, 0.48]		27.87 - 33.26
	α_d	0.22	[0.11, 0.33]		11.98 - 17.27
	T_d	0.21	[0.10, 0.31]		3.34 - 8.83
	silt	-0.03*	[-0.16, 0.09]		6.81 - 11.62

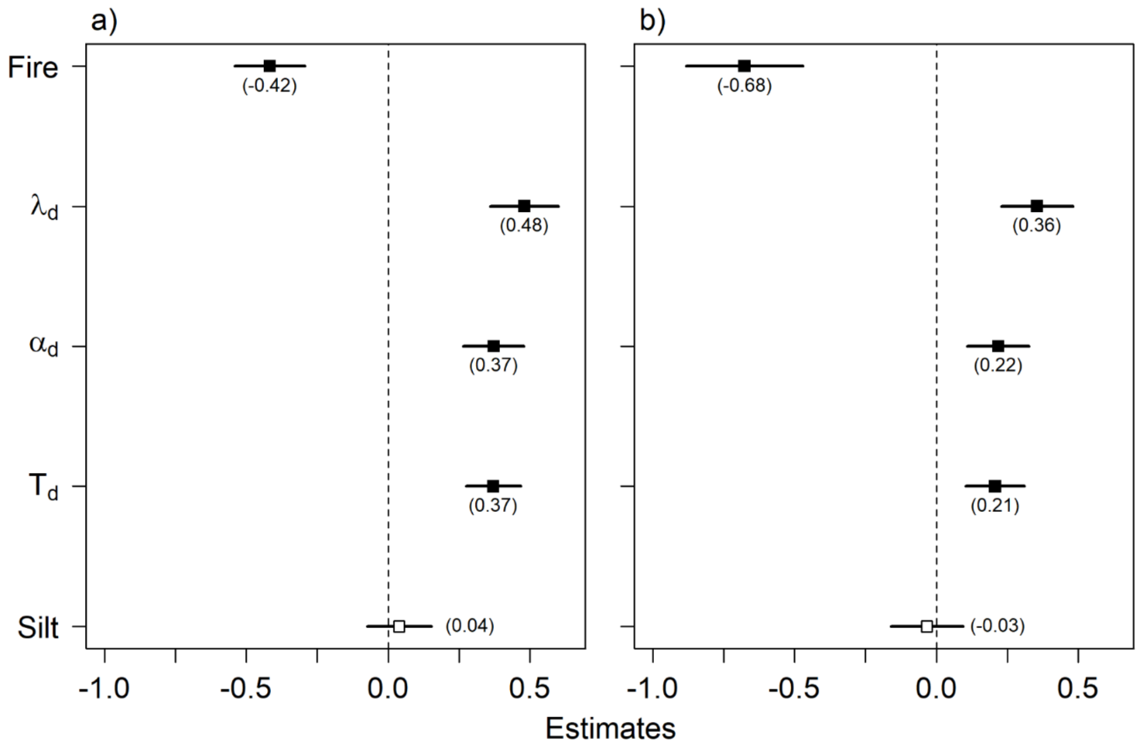

Figure S10. Effect of each predictor variable on a) tree cover and b) $PAVD_{max}$. Predictor variables were standardized such that their GLM coefficient magnitude is a measure of their importance in the model. The median estimate (values in parentheses) and 95% confidence interval (error bars) are based on 1000 GLMs (see Section 2.5 and Supporting information S4). Terms are not significant (open symbol) when the confidence interval includes zero (dashed vertical line). Predictor variables are: fire frequency (fire), mean daily precipitation frequency in dry season (λ_d), mean daily precipitation intensity in dry season (α_d), mean length of dry season (T_d), soil silt content (silt).

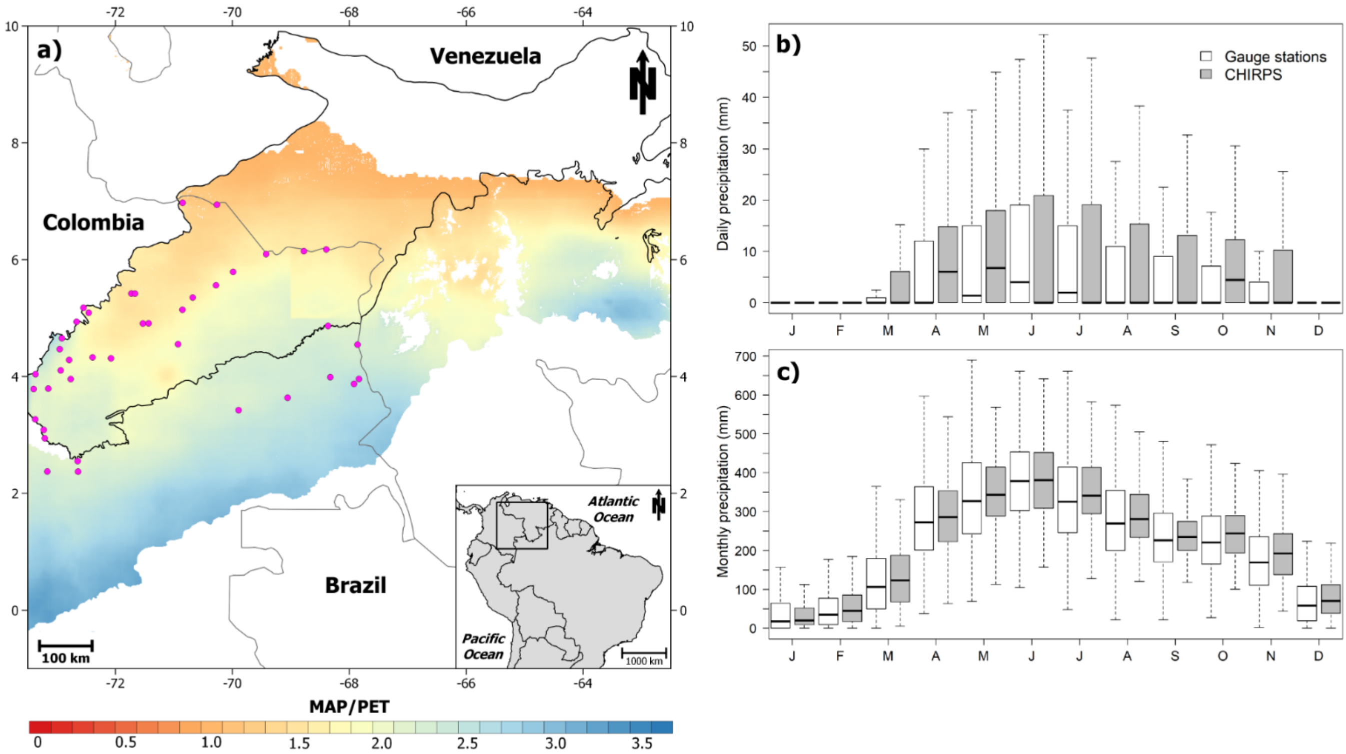

Figure S11. a) Precipitation gauge stations (magenta points, N=40) from the Colombian Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) within the study area with ≤10% of daily missing data for the period 2004-2015 (Valencia and others 2023). Daily b) and c) monthly precipitation from gauge stations and the corresponding CHIRPS pixels for the period 2004-2015; months with more than 5 days of missing values in gauge stations were excluded from the analysis. In b) and c), outlier values were excluded for visualization purposes.

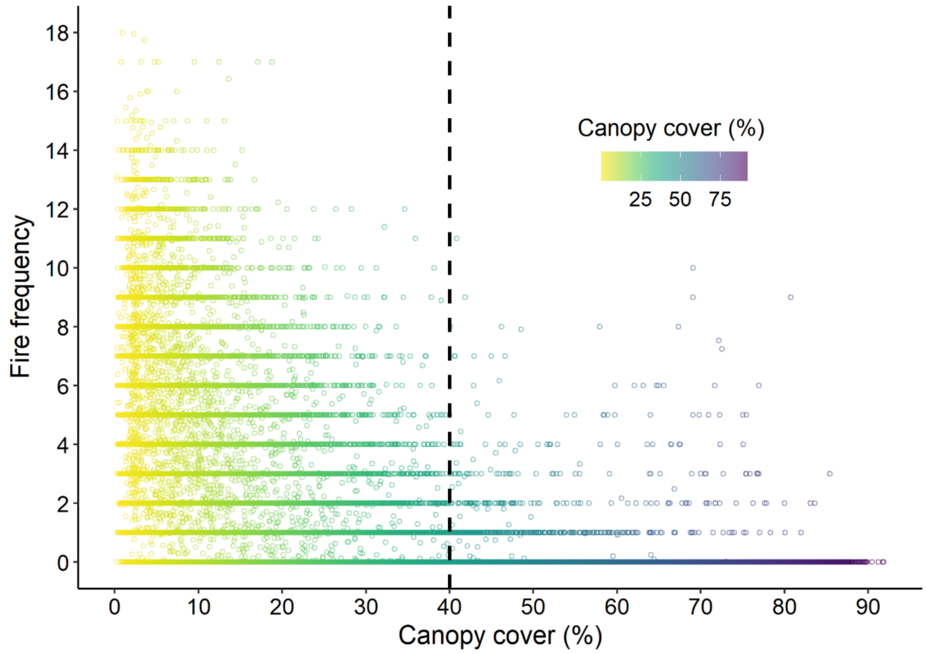

Figure S12. Relationship between canopy cover and fire frequency for forest and savanna pixels that occur in the same MAP (2360 to 3070 mm) and λ_d (0.21 to 0.43) range (Figure 4b) for the period 2001-2019. The vertical dashed line indicates the canopy cover threshold used to define savanna (< 40 %) and forest (≥ 40%) pixels.

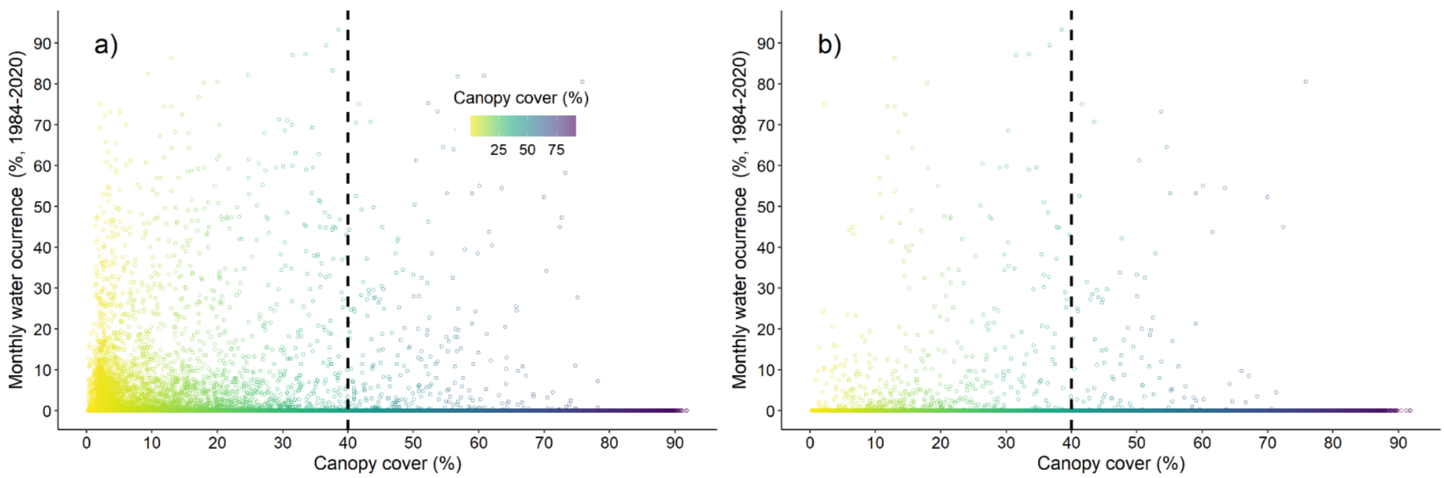

Figure S13. Relationship between canopy cover and percentage of monthly water occurrence during the period 1984-2020 obtained from Pekel and others (2016). a) all pixels within our study area (N = 201474) and b) forest and savanna pixels (N=68677) in the same climatic space based on MAP (2360 to 3070 mm) and λ_d (0.21 to 0.42). Monthly water occurrence was resampled from 30 m to 0.01° (~1 km).

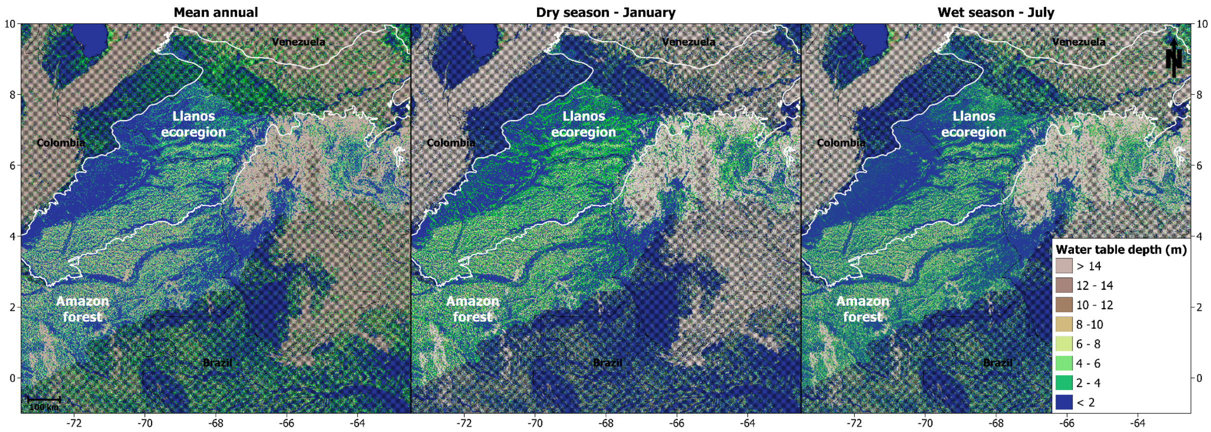

Figure S14. Water table depth (WT) for our study area (highlighted pixels) in northern South America at 0.01° resolution from Fan et al. (2013). The white line indicates the Llanos ecoregion boundary. Pixels outside the study area were excluded following the criteria described in Study Area Delimitation and Forest-Savanna Discrimination Section.

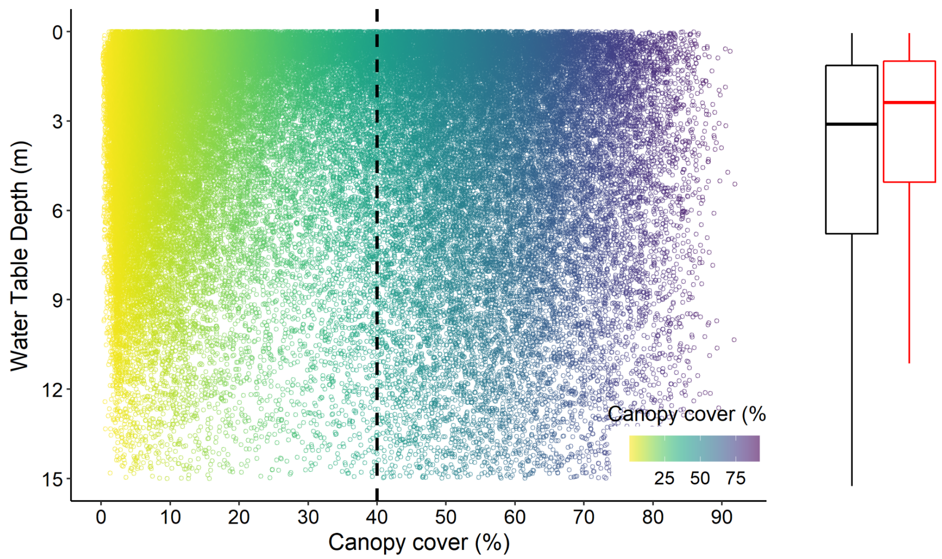

Figure S15. Relationship between canopy cover and water table depth for forest and savanna pixels (N=68677) in the same climatic space based on MAP (2360 to 3070 mm) and λ_d (0.21 to 0.42). Boxplots show the values of water table depth for savanna (red) and forest (black) pixels defined by canopy cover.

References

Adzhar R, Kelley DI, Dong N, Torello Raventos M, Veenendaal E, Feldpausch TR, Philips OL, Lewis S, Sonké B, Taedoumg H, Schwantes Marimon B, Domingues T, Arroyo L, Djagbletey G, Saiz G, Gerard F. 2021. Assessing MODIS Vegetation Continuous Fields tree cover product(collection 6): performance and applicability in tropical forests and savannas. Biogeochemistry: Land https://bg.copernicus.org/preprints/bg-2020-460/bg-2020-460.pdf. Last accessed 10/05/2022

Bernardi RE, Staal A, Xu C, Scheffer M, Holmgren M. 2019. Livestock Herbivory Shapes Fire Regimes and Vegetation Structure Across the Global Tropics. Ecosystems

22:1457–65.

Calders K, Armston J, Newnham G, Herold M, Goodwin N. 2014. Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR. Agric For Meteorol 194:104–17.

Compagnoni A, Levin S, Childs DZ, Harpole S, Paniw M, Römer G, Burns JH, Che-Castaldo J, Rüger N, Kunstler G, Bennett JM, Archer CR, Jones OR, Salguero-Gómez R, Knight TM. 2021. Herbaceous perennial plants with short generation time have stronger responses to climate anomalies than those with longer generation time. Nat Commun 12:1824.

Decuyper M, Mulatu KA, Brede B, Calders K, Armston J, Rozendaal DMA, Mora B, Clevers JGPW, Kooistra L, Herold M, Bongers F. 2018. Assessing the structural differences between tropical forest types using Terrestrial Laser Scanning. For Ecol Manag 429:327–35.

DiMiceli C, Townshend J, Carroll M, Sohlberg R. 2021. Evolution of the representation of global vegetation by vegetation continuous fields. Remote Sens Environ 254:112271.

D'Onofrio D, von Hardenberg J, Baudena M. 2018. Not only trees: Grasses determine African tropical biome distributions via water limitation and fire. Glob Ecol Biogeogr 27:714–25.

D'Onofrio D, Sweeney L, von Hardenberg J, Baudena M. 2019. Grass and tree cover responses to intra-seasonal rainfall variability vary along a rainfall gradient in African tropical grassy biomes. Sci Rep 9:2334.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46.

Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S, Armston J, Tang H, Duncanson L, Hancock S, Jantz P, Marselis S, Patterson PL, Qi W, Silva C. 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci Remote Sens 1:100002.

Hirota M, Holmgren M, Van Nes EH, Scheffer M. 2011. Global Resilience of Tropical Forest and Savanna to Critical Transitions. Science 334:232–5.

Hoyos N, Correa‐Metrio A, Jaramillo C, Villegas JC, Escobar J. 2022. Effects of consecutive dry and wet days on the forest–savanna boundary in north‐west South America. Gillespie T, editor. Glob Ecol Biogeogr 31:347–61.

Li X, Wessels K, Armston J, Hancock S, Mathieu R, Main R, Naidoo L, Erasmus B, Scholes R. 2023. First validation of GEDI canopy heights in African savannas. Remote Sens Environ 285:113402.

Luthcke S, Sabaka T, Nicholas J, Preaux S, Hofton M. (2021). GEDI L3 Gridded Land Surface Metrics, Version 1. *ORNL DAAC*.

Meeussen C, Govaert S, Vanneste T, Calders K, Bollmann K, Brunet J, Cousins SAO, Diekmann M, Graae BJ, Hedwall P-O, Krishna Moorthy SM, Iacopetti G, Lenoir J, Lindmo S, Orczewska A, Ponette Q, Plue J, Selvi F, Spicher F, Tolosano M, Verbeeck

H, Verheyen K, Vangansbeke P, De Frenne P. 2020. Structural variation of forest edges across Europe. For Ecol Manag 462:117929.

Pekel J-F, Cottam A, Gorelick N, Belward AS. 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540:418–22.

Rishmawi K, Huang C, Zhan X. 2021. Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data. Remote Sens 13:442.

Silva CA, Hamamura C, Valbuena R, Hancock S, Cardil A, Broadbent E, N., Almeida, D.R.A., Silva Junior C.H.L, Klauberg C. rGEDI: NASA's Global Ecosystem Dynamics Investigation (GEDI) Data Visualization and Processing. version 0.1.9, accessed on October. 22 2020, available at: https://CRAN.R-project.org/package=rGEDI

Staal A, Dekker SC, Xu C, van Nes EH. 2016. Bistability, Spatial Interaction, and the Distribution of Tropical Forests and Savannas. Ecosystems 19:1080–91.

Tang H, Armston J, Dubayah R. (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics. *Goddard Space Flight Center: Greenbelt, MD, USA*.

Valencia S, Marín DE, Gómez D, Hoyos N, Salazar JF, Villegas JC. 2023. Spatio-temporal assessment of Gridded precipitation products across topographic and climatic gradients of Colombia. Atmospheric Res:106643.