Supplementary Information for

Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate

C. A. Quesada[1,2], O. L. Phillips[1], M. Schwarz[3], C. I. Czimczik[4], T. R. Baker[1], S. Patiño[1,4,†], N. M. Fyllas[1], M. G. Hodnett[5], R. Herrera[6], S. Almeida[7,†], E. Alvarez D´avila[8], A. Arneth[9], L. Arroyo[10], K. J. Chao[1], N. Dezzeo[6], T. Erwin[11], A. di Fiore[12], N. Higuchi[2], E. Honorio Coronado[13], E. M. Jimenez[14], T. Killeen[15], A. T. Lezama[16], G. Lloyd[17], G. López-González[1], F. J. Luizão[2], Y. Malhi[18], A. Monteagudo[19,20], D. A. Neill[21], P. Núñez Vargas[19], R. Paiva[2], J. Peacock[1], M. C. Peñuela[14], A. Peña Cruz[20], N. Pitman[22], N. Priante Filho[23], A. Prieto[24], H. Ramírez[16], A. Rudas[24], R. Salomão[7], A. J. B. Santos[2,25,†] , J. Schmerler[4], N. Silva[26], M. Silveira[27], R. Vásquez[20], I. Vieira[7], J. Terborgh[22], and J. Lloyd[1,28]

1School of Geography, University of Leeds, LS2 9JT, UK
2Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil
3Ecoservices, 07743 Jena, Germany
4Max-Planck-Institut fuer Biogeochemie, Jena, Germany
5Centre for Ecology and Hydrology, Wallingford, UK
6Instituto Venezolano de Investigaciones Científicas, Caracas Venezuela
7Museu Paraense Emilio Goeldi, Belém, Brazil
8Jardin Botanico de Medellin, Medellin, Colombia
9Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Garmisch-Partenkirchen, Germany
10Museo Noel Kempff Mercado, Santa Cruz, Bolivia
11Smithsonian Institution, Washington, DC 20560-0166, USA
12Department of Anthropology, New York University, New York, NY 10003, USA
13IIAP, Apartado Postal 784, Iquitos, Peru
14Universidad Nacional de Colombia, Leticia, Colombia
15Centre for Applied Biodiversity Science, Conservation International, Washington DC, USA
16Faculdad de Ciencias Forestales y Ambientales, Univ. de Los Andes, Merida, Venezuela
17Integer Wealth Management, Camberwell, Australia
18School of Geography and the Environment, University of Oxford, Oxford, UK
19Herbario Vargas, Universidad Nacional San Antonio Abad del Cusco, Cusco, Peru
20Proyecto Flora del Perú, Jardin Botanico de Missouri, Oxapampa, Peru
21Herbario Nacional del Ecuador, Quito, Ecuador
22Centre for Tropical Conservation, Duke University, Durham, USA
23Depto de Fisica, Universidade Federal do Mato Grosso, Cuiabá, Brazil
24Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá, Colombia
25Depto de Ecologia, Universidade de Brasilia, DF, Brazil
26Empresa Brasileira de Pesquisas Agropecuárias, Belém, Brazil
27Depto de Ciências da Natureza, Universidade Federal do Acre, Rio Branco, Brazil
28Centre for Tropical Environmental and Sustainability Science (TESS) and School of Earth and Environmental Studies, James Cook University, Cairns, Queensland 4878, Australia
†deceased

A: Separation of stand-level parameters into underlying components

The above ground biomass of a stand per unit area , B, can be expressed as

$$B = \frac{\sum_{t=1}^{N} \xi A_t \rho_{w(t)}}{a} \quad , \qquad \text{(S1)}$$

where A_t is the basal area (m^2) of tree t , $\rho_{w(t)}$ is the wood density of tree t (kg m^{-3}) and a is the plot area (ha) with ξ being an allometric coefficient transforming A_t into a wood volume estimate (V_t) and N is the number of trees measured. In our analysis, ξ is taken as a constant (being based on Chambers et al., 2001) and so, being normalised by the area sampled, differences in B must be

attributable to variations in the average values of A_t, $\varrho_{w(t)}$ or in N/a, the latter being equal to S, the stem density (# ha^{-1}).

Separating Eq. 1 into its components: The basal area of the stand A_t (m^2 ha^{-1}) can be expressed as

$$\frac{A_B = \sum_{t=1}^{N} A_t}{a} \quad , \qquad \text{(S2)}$$

and the mean wood density of the stand, ϱ_W, can be expressed as

$$\frac{\rho_W = \sum_{sp=1}^{j} \rho_{w(sp)} \sum_{t=1}^{k(sp)} \xi A_{t(sp)}}{a} \quad , \qquad \text{(S3)}$$

where $\varrho_{w(sp)}$ is the characteristic wood density for each species (*sp*) and *k(sp)* is the number of each species in the plot. Mean stand wood density so calculated is therefore a volume weighted estimate, and from which it follows that

$$B = \xi A_B \rho_W \quad , \qquad \text{(S4)}$$

with, after the inclusion of the appropriate corrections (Malhi et al., 2004; Lewis et al., 2004), basal area growth rates (G_B) being related to the basal area growth rates of individual trees (G_t), and with above ground wood productivity (W_P) similarly being related to G_B (to good approximation) as

$$W_P = \xi G_B \rho_W \quad . \qquad \text{(S5)}$$

Likewise, the stand level turnover rate (φ) can be expressed and

$$\varphi = \frac{0.5(r+m)}{S} \qquad \text{(S6)}$$

where r is the mean rate of recruitment (trees ≥ 10 cm dbh observed for the first time) expressed as stems ha^{-1} yr^{-1} and m is the mortality rate (including standing dead trees noted as such for the first time), also in the same units. Thus, φ is effectively a probability with between stand differences attributable to variations attributable to either (r+m) or S. Similar to the above, turnover rate is also potentially expressible in biomass units as the mean of W_P and the biomass mortality rate, M; this being the estimated rate of loss from the living biomass pool expressed as t ha^{-1} yr^{-1} calculated according to the same principles as W_P above.

Investigating first the importance of plot-to-plot variations in S, Fig. S1 shows (top row) A_t and A_B vary with S and with each other. Across our dataset, S varies by a factor of just over 3, with the lowest S being a 'bamboo forest' in Acre Brazil (DOI-02) on a plinthosol and the two highest values being for 'white sand forests' (SCR-04 and ZAR-01), both on podzols. There is a significant decline in A_t with S (R = -0.51, $p < 0.001$) and with a log-log (power) scaling factor (β) of -1.40. There is no significant relationship between A_B and S (p = 0.08), so even though the relationship between A_t and A_B is statistically strong (R = 0.71, $p < 0.001$) β is estimated at only 0.88. This

means that, although most of the variation in A_B is due to differences in A_t, effects of A_t on A_B have a scaling factor of less than 1.00, this reflecting a 'self-thinning' tendency.

A similar analysis is undertaken for tree turnover rates and S in the middle three panels of Fig. S1. This clearly shows that turnover variations are not related to variations in S and with $\beta = 1.01$ in the middle right panel indicating that variations in φ accurately reflect differences in the number of stems turning over per unit area ($R = 0.92$, $p < 0.001$) rather than the different S.

In the bottom three panels of Fig. S1, G_t and G_B covariation with S is examined. This shows that there is a strong tendency for G_t to decline with S ($R = 0.92$, $p < 0.001$) and with $\beta = -1.91$ which results in no relationship between S and G_B and a strong relationship between G_t and G_B but with a β of only 0.82.
Thus we may conclude, similar to the analysis of basal area stocks above, that stand-to-stand variation in basal area growth rates reflects to a large degree variations in the growth rates of individual trees. But that this is not a 1:1 relationship because of G_t systematically declining with increasing S. Of additional interest, (middle low panel) is that stands on typically more fertile soils such as the cambisols tend to sit above the line of best fit, and with those on relatively infertile soils such as ferralsols sitting below the fitted line. This means that although G_t tends to decline with increasing S independent of soil type, at any given S trees growing on the more fertile soil types tend to have a higher growth rate. It is also for this reason that the plots segregate according to soil type along the G_B; G_t line of best fit in the last panel.

This analysis is further developed in Fig. S2, with the importance of plot-to-plot variation in ϱ_W investigated. Taking first the top row of panels, variations in A_B and B with ϱ_W and with each other are shown. This shows that variations in B mostly arise as a consequence of differences in A_B ($R = 0.82$, $p < 0.001$; $\beta = 1.15$) rather than differences in ϱ_W, especially a low values of A_B. That is to say, as is also clear from the middle top panel ($R = 0.34$, $p < 0.01$; $\beta = 2.34$) the highest biomass stands only occur in plots which also have a high ϱ_W and with there being considerable variation in B at high A_B due to differences in ϱ_W. But not at lower A_B.

Probability and biomass based turnover estimates are compared with ϱ_W and with each other in the middle row panels of Fig. S2. This shows that ϱ_W turns out to be a particularly good predictor of φ ($R = -0.63$ $p < 0.001$; $\beta = -4.2$), though as mentioned in the discussion of the main paper, this relationship is not necessarily causative. Because of this negative φ; ϱ_W correlation, β is much less at = 3.0 for the biomass based turnover metric and with a lesser level of significance ($R = -0.41$; $p < 0.005$) but with the relationships between the two turnover measures whilst still being strongly significant ($R = -0.59$; $p < 0.001$) also having a β of only 0.72.

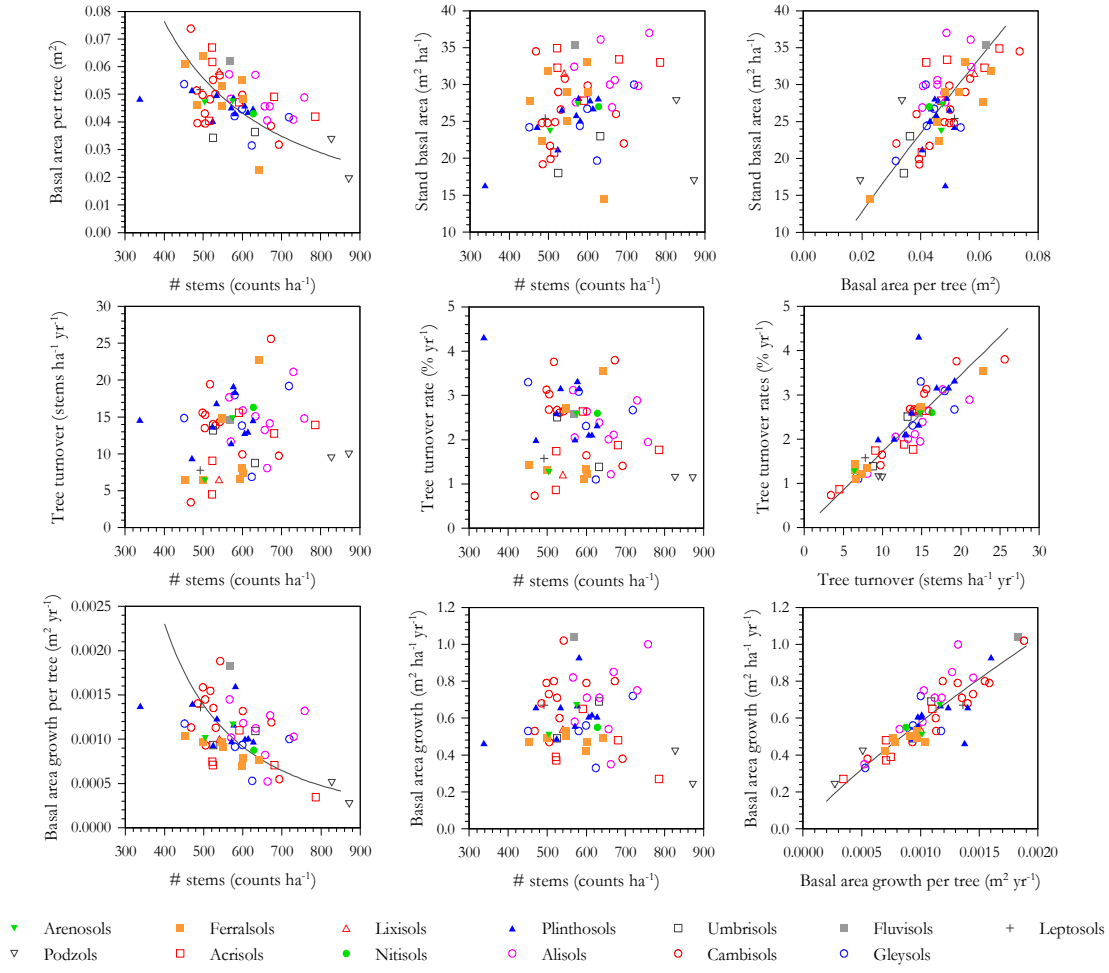

Figure S1. Relationship between stem density and several metrics of stand dynamics. Where significant with normal least squares regression, standard major axis scaling relationships are also shown, *viz*: $y = ax^{\beta}$ with the appropriate values of β reported in the text

Finally, the relationships between G_B, W_P and ϱ_W are examined in the lower three panels of Fig. S2. This shows, as for φ, a strong decline in G_B with increasing ϱ_W (R = -0.64 $p < 0.001$; β = -3.2) with the relationship between W_P and ϱ_W less dramatic (R = -0.42 $p < 0.006$; β = -3.2). It is also clear that much of the Basin-wide variation in W_P is due to differences in G_B (bottom right panel; R = 0.81 $p < 0.001$; β = 0.89).

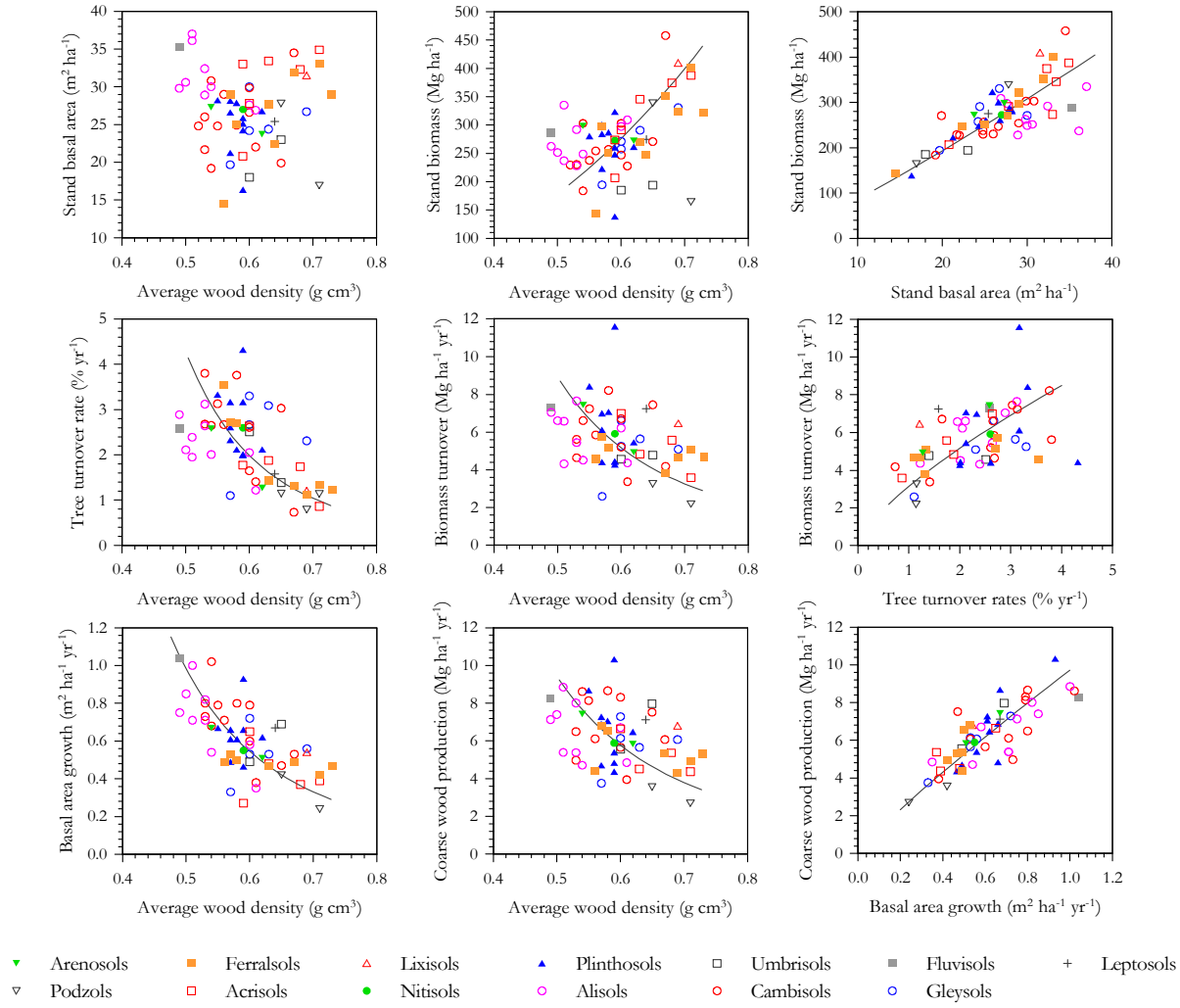

Figure S2. Relationship between stand wood density and several metrics of stand dynamics. Where significant with normal least squares regression, standard major axis scaling relationships are also shown, *viz*: $y = ax^{\beta}$ with the appropriate values of β reported in the text

B: Relationships of stand basal area and stem density with their geographical space with subsequent analysis of their underlying climatic and edaphic controls

Relationship among stand basal area, number of trees per hectare and the geographical space

Figure S3 shows the geographical distribution of forest stand basal area (A_B) and number of stems per hectare (S). The highest forest basal area is found at eastern and northern parts of Amazon as well as at some Ecuadorian forests, while western and southern portions of Amazonia have smaller A_B, as has been previously reported by Malhi et al. (2006). By contrast S (which ranges from 338 to 872 trees >10cm DBH ha^{-1}) did not show a clear geographical pattern.

Although showing some level of correlation with latitude and longitude, both A_B and S seem to have low level of spatial autocorrelation (Fig S4). To account for spatial autocorrelation, eigenvector-based spatial filters (Bocard and Legendre, 2002) was calculated for A_B, and S and used in a multiple regression framework as additional predictors. Following the SEVM-1 procedure, six filters were selected for A_B (filters 1,2, 3, 4, 5 and 8). Filter selection for the SEVM-2 procedure selected for three filters (2, 3 and 5) and the SEVM-3 procedure only retained one filter for AB (filter 5). The SEVM-1 procedure selected eight filters for S (1 to 8) while SEVM-2 selected three (filter 1, 5, 8). SEVM-3 did not select for any eigenvector filter in S. Spatial structures alone could explain 0.56 and 0.24 of their variation (for A_B and S, respectively).

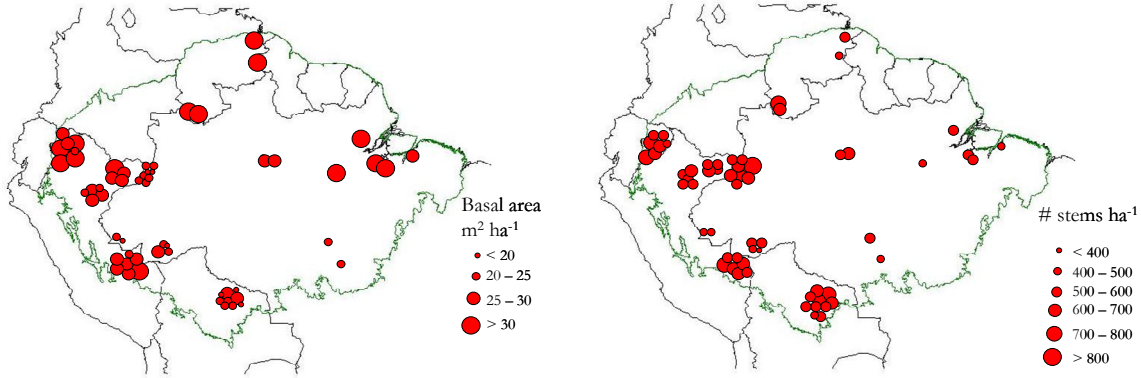

Figure S3. Geographical distribution of forest basal area (A_B) and number of stems (S), across Amazonia. Size of circles represents variations among sites, see legend for details.

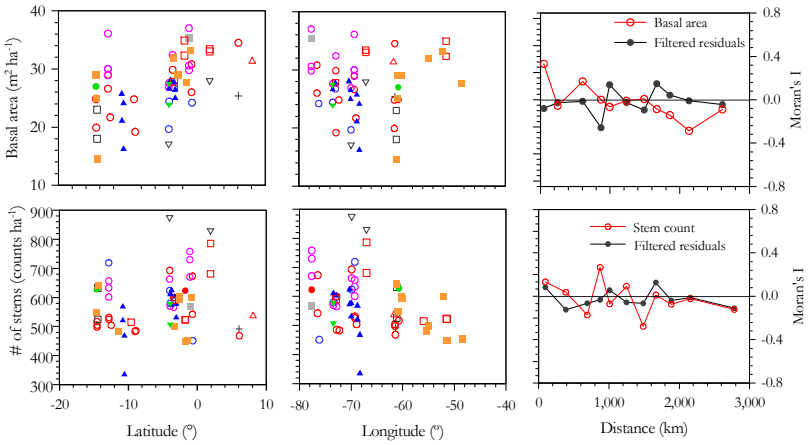

Figure S4. Correlations between forest basal area and stem density with the geographic space. Moran's *I* correlograms are also given showing spatial autocorrelation for $\boldsymbol{A_B}$ but with spatial filters being able to effectively remove its effect from regression residuals. Moran I correlogram for *S* shows little spatial autocorrelation.

Edaphic and climatic controls on stand basal area

In order to investigate relationships between A_B and environmental factors, spatially adjusted Spearman correlations were performed for a total of 28 predictors (Table S1). Associated p values and degrees of freedom were adjusted to account for spatial autocorrelation using the Dutilleul method (Dutilleul, 1993). Soil chemistry predictors were generally inversely related to A_B which may suggest that lower A_B may occur on more fertile soils. The relationships between soil fertility parameters and A_B were however relatively modest (Fig. S5), and no significant correlation was observed (Table S1). Similarly, soil physical properties also showed weak relationships with A_B, with all soil physical parameters being negatively correlated to it (i.e. worse physical condition resulting in lower A_B), with the exception of topography which showed a positive correlation. Correlations with all soil physical predictors turned out as not significant, but the combined indexes of physical properties (Π_1 and Π_2) showed some indication that lower A_B may be associated to adverse soil physical conditions, at least in a large proportion of the study sites (Fig. S6). It seems however that A_B is more strongly related to climatic factors than to soil properties. Figure S7 shows strong relationships between A_B and mean annual temperature, average annual precipitation and minimum dry season precipitation, suggesting effect of both amount and distribution of rainfall. Correlations between these variables and forest A_B were of significance (Table S1), with the dry season precipitation emerging as the stronger factor influencing forest basal area.

We have then performed a multi model selection based on AIC including all variables in this study. The best OLS model fit for A_B included Π_2, $[P]_T$, $[K]_E$, and P_A as predictors which were used into a multiple regression (Table. S2). The regression model resulted highly significant (p<0.001) and explaining about to 42% of variation in A_B. All predictors turned out significant (p=0.006, p<0.001,

p=0.019 and p<0.001, respectively). Additional models within the ΔAIC<2 interval have also selected for Π_2, $[P]_T$ and P_A, but with varying cation measures instead or in addition to $[K]_E$.

Model selection with the inclusion of SEVM-1 spatial filters resulted in P_A and Π_1 as predictors for A_B, with only P_A resuming significant (p=0.014 and p=0.109, respectively). Similarly, once SEVM-1 filters were applied into the OLS model, only P_A retained its significant (Table S2). Performing model selection with the more conservative SEVM-2 group of filters resulted in a best model containing $[P]_T$, P_A, T_A and Σ_B, suggesting that both climate and soil properties may be significant predictors of A_B. All variables resumed significant with SEVM-2 filters (p=0.031, p<0.001 and p=0.045 for $[P]_T$, P_A, Σ_B, respectively), but with T_A being only marginally significant at p=0.083. The inclusion of SEVM-2 filters into OLS regression also changed substantially the significance levels, with only P_A remaining highly significant (p<0.001) but with $[P]_T$ being marginally significant (p=0.066).

The regression residuals from the OLS model were then tested for correlation against nine different eigenvector filters (SEVM-3), with this resulting in one filter (filter 5) being selected to inclusion in the OLS regression model. Addition of SEVM-3 filter to the OLS regression only slightly changed the significance levels, with Π_2, $[P]_T$, $[K]_E$, and P_A with all predictors still being significant after the addition of SEVM-3 (p=0.011, p=0.005, p=0.014 and p<0.001, respectively).

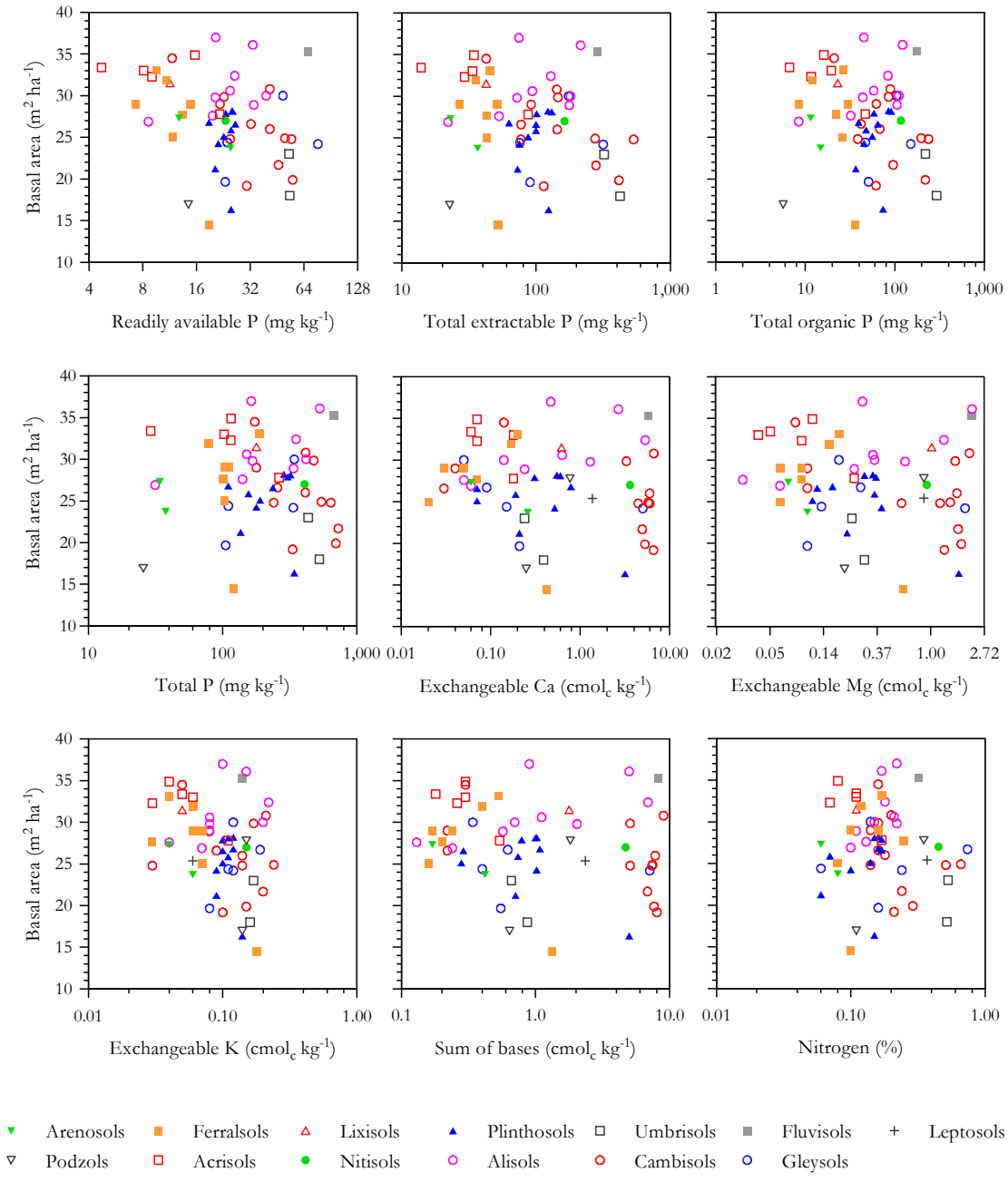

Figure S5. **Relationships between stand basal area and different soil fertility parameters**

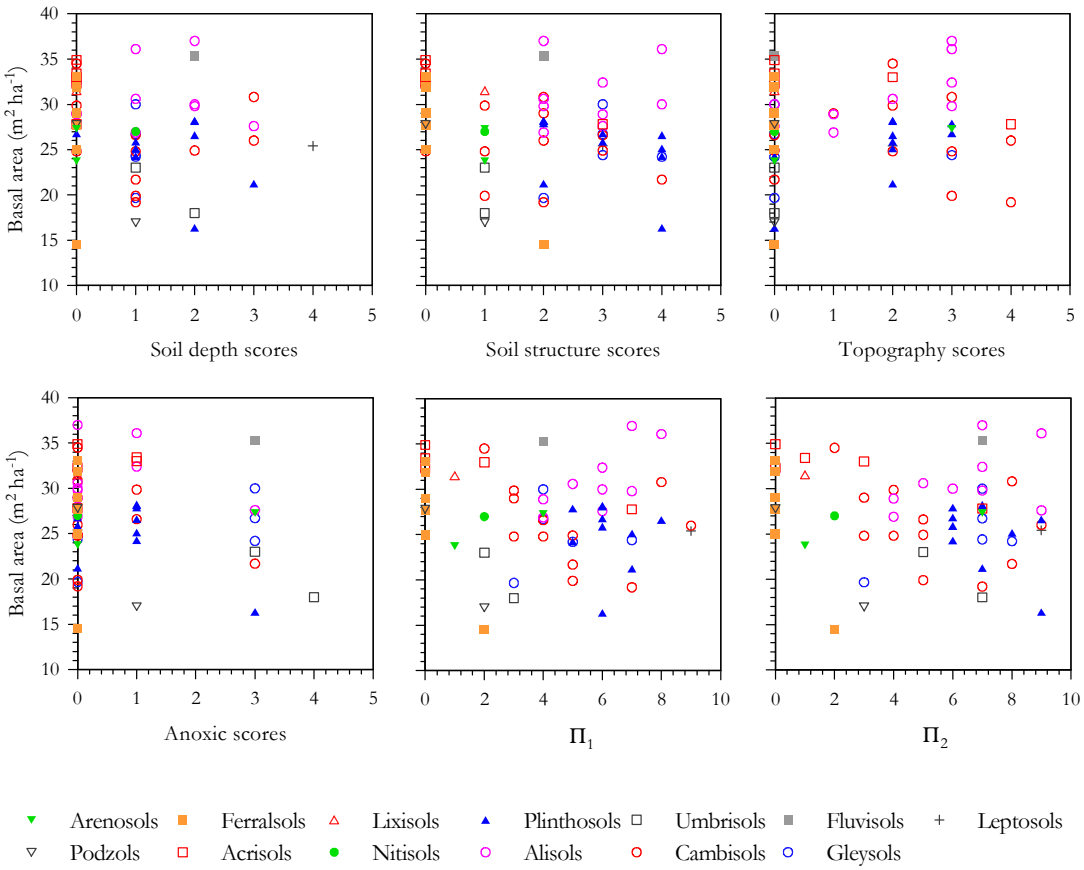

Figure S6. Relationships between stand basal area and soil physical properties

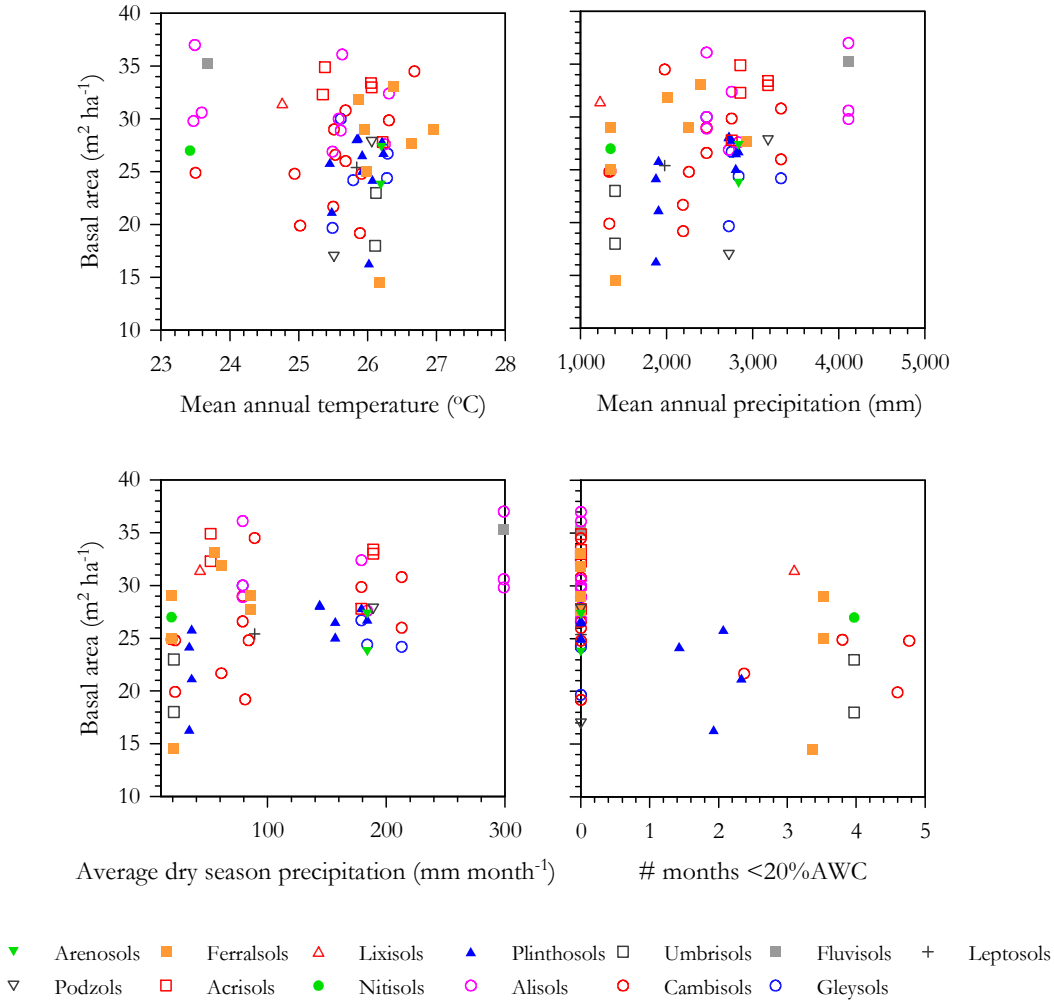

Figure S7. Relationships between stand basal area and climatic factors

Table S1 Spearman correlations between forest basal area and different predictors, adjusted for spatial autocorrelation by Dutillieul's method (Dutillieul, 1993).

Variable	Spearman rho	p value	p value adj	df adj
Sand	-0.087	0.489	0.529	41.42
Clay	0.185	0.139	0.193	38.70
Silt	-0.060	0.633	0.696	33.65
Soil depth	-0.232	0.218	0.199	54.43
Soil structure	-0.295	0.145	0.305	25.00
Topography	0.047	0.186	0.244	38.90
Anoxic	-0.071	0.117	0.114	50.62
Π_1	-0.110	0.614	0.671	35.66
Π_2	-0.164	0.287	0.376	34.76
Temperature	-0.051	0.194	0.196	49.34
Precipitation	**0.370**	**<0.001**	**0.114**	**12.00**
Dry s. min. precip	**0.333**	**0.013**	**0.131**	**18.73**
# month <20% AWC	**0.318**	**<0.001**	**0.046**	**15.61**
pH	-0.113	0.272	0.205	66.29
P_{TE}	-0.199	0.082	0.155	33.53
$P_{readily\ available}$	-0.277	0.141	0.121	55.37
P_{inorg}	-0.196	0.097	0.150	37.70
P_{org}	-0.174	0.099	0.186	32.37
P_{total}	-0.017	0.642	0.694	35.78
Nitrogen	0.032	0.281	0.342	38.97
Carbon	-0.025	0.197	0.321	29.85
C:N	-0.028	0.787	0.826	33.46
Ca	-0.116	0.800	0.606	204.88
Mg	-0.070	0.984	0.977	102.28
K	-0.284	0.046	0.057	45.84
Al	0.172	0.304	0.115	117.25
Sum of bases	-0.108	0.786	0.604	181.01
ECEC	-0.003	0.760	0.730	63.56

Table S2. Lowest AIC model fits for the prediction of G_B, S and A_B with and without the use of spatial filters. First column give results for the OLS regression, with the second column giving results for the OLS + SEVM-1 (in brackets) and results for SEVM-1 in bold. Third column show results for OLS+SEVM-2 (in brackets) and SEVM-2 in bold. The fourth column shows results for OLS+SEVM-3. The upper line gives first the standardized coefficients (β) and their level of significance (p). Regression models where no filters or predictors have been selected are denoted with a "--- ".

	OLS		SEVM-1		SEVM-2		SEVM-3	
G_B model	β	p	β	p	β	p	β	p
$Log[P]_T$	0.432	<0.001	(0.288) **0.337**	(0.035) **0.003**	(0.262) **0.315**	(0.040) **0.003**	0.432	<0.001
T_A	-0.252	0.022	(-0.180) **--**	(0.227) **--**	(-0.162) --	(0.210) --	-0.252	0.022
P_A	0.375	0.002	(-0.035) **--**	(0.838) **--**	(0.224) **--**	(0.063) **--**	0.375	0.002
Π_1	0.215	0.077	(0.077) **0.519**	(0.678) **0.014**	(0.144) **--**	(0.226) **--**	0.215	0.077
R_a	--	--	--	--	-- **-0.298**	-- **0.004**	--	--
Filter 1	--	--	(0.311) **0.268**	(0.098) **0.198**	--	--	--	--
Filter 2	--	--	(0.014) **0.011**	(0.923) **<0.001**	--	--	--	--
Filter 3	--	--	(-0.336) **-0.456**	(0.013) **<0.001**	(-0.323) **-0.450**	(0.015) **<0.001**	--	--
Filter 4	--	--	(0.285) **0.235**	(0.015) **0.635**	(0.215) **0.175**	(0.034) **0.057**	--	--
Filter 5	--	--	(-0.264) **-0.229**	(0.023) **0.014**	(-0.170) **-0.355**	(0.108) **<0.001**	--	--
Filter 6	--	--	(-0.121) **-0.159**	(0.195) **0.788**	--	--	--	--
AIC	-61.59		(-64.00) **-70.95**		(-67.21) **-72.74**		-61.59	

	OLS		SEVM-1		SEVM-2		SEVM-3	
S model	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*
Π_1	--	--	--	--	--	--	--	--
					-0.045	**0.790**		
Log[Ca]e	--	--	--	--	--	--	--	--
			-0.406	**0.005**	--	--		
$Log[K]_e$	-0.584	<0.001	(0.519)	(0.005)	(0.555)	(<0.001)	-0.584	<0.001
			--	**--**	**0.379**	**0.011**		
$Log[Mg]_e$	-0.494	0.002	(-0.393)	(0.039)	(-0.438)	(0.012)	-0.494	0.002
			0.411	**0.011**				
$Log[P]_a$	-0.322	0.024	(-0.280)	(0.132)	(-0.340)	(0.024)	-0.322	0.024
			--	**--**	**-0.421**	**0.011**		
T_A	-0.330	0.004	(-0.214)	(0.200)	(-0.306)	(0.033)	-0.330	0.004
			--	**--**	**--**	**--**		
P_D	0.364	<0.001	(1.006)	(0.090)	(0.393)	(0.214)	0.364	<0.001
			1.718	**<0.001**	**0.628**	**0.023**		
Filter 1	--	--	(0.584)	(0.249)	(0.070)	(0.809)	--	--
			1.142	**0.007**	**0.258**	**0.339**		
Filter 2	--	--	(0.078)	(0.530)	--	--	--	--
			0.085	**0.494**				
Filter 3	--	--	(-0.256)	(0.210)	--	--	--	--
			-0.507	**0.006**				
Filter 4	--	--	(0.066)	(0.675)	--	--	--	--
			0.189	**0.157**				
Filter 5	--	--	(-0.092)	(0.611)	(0.059)	(0.651)	--	--
			-0.226	**0.186**	**0.047**	**0.733**		
Filter 6	--	--	(-0.009)	(0.939)	--	--	--	--
			0.015	**0.895**				
Filter 7	--	--	(-0.005)	(0.963)	--	--	--	--
			0.013	**0.911**	**--**	**--**		
Filter 8	--	--	(-0.047)	(0.792)	(0.091)	(0.507)	--	--
			-0.227	**0.150**	**0.138**	**0.283**		
AIC	677.27		(698.20)		(684.03)		677.27	
			694.04		**691.76**			

	OLS		SEVM-1		SEVM-2		SEVM-3	
A_B model	β	p	β	p	β	p	β	p
$\text{Log}[P]_T$	0.493	<0.001	(0.265) **--**	(0.141) **--**	(0.268) **0.305**	(0.066) **0.031**	0.404	0.005
$\text{Log}[K]_e$	-0.334	0.019	(-0.215) **--**	(0.138) **--**	(-0.206) --	(0.131) --	-0.336	0.014
Π_2	-0.424	0.006	(-0.008) **--**	(0.970) **--**	(-0.057) **--**	(0.760) **--**	-0.383	0.011
P_A	0.650	<0.001	(0.560) **0.519**	(0.010) **0.014**	(0.450) **0.399**	(<0.001) **<0.001**	0.575	<0.001
T_A	--	--	--	--	-- **-0.194**	-- **0.083**	--	--
Σ_B	--	--	--	--	-- **-0.284**	-- **0.045**	--	--
Π_1	--	--	-- **0.256**	-- **0.109**	--	--	--	--
Filter 1	--	--	(0.117) **0.245**	(0.559) **0.198**	--	--	--	--
Filter 2	--	--	(0.360) **0.556**	(0.030) **<0.001**	(0.320) **0.373**	(0.028) **<0.001**	--	--
Filter 3	--	--	(-0.287) **-0.395**	(0.029) **<0.001**	(-0.290) **-0.384**	(0.022) **<0.001**	--	--
Filter 4	--	--	(0.077) **0.063**	(0.582) **0.635**	--	--	--	--
Filter 5	--	--	(0.239) **0.272**	(0.041) **0.014**	(0.262) **0.271**	(0.015) **0.010**	0.254	0.026
Filter 8	--	--	(-0.023) **0.027**	(0.835) **0.788**	--	--	--	--
AIC	292.40		(295.50) **290.13**		(286.26) **282.611**		289.39	

Edaphic and climatic controls on stem density

Correlations between environmental factors and *S* are shown on Table S3. Similar to A_B, bivariate plots of soil fertility parameters show little capacity to explain variations in *S* (Fig S8). Most fertility parameters did not resulted significantly correlated to *S*, exception being exchangeable Ca and Mg as well as the sum of bases, which showed weak but significant correlations (ρ -0.119, -0.112 and -0.085, respectively). There is however a tendency to soil fertility to be negatively correlated to *S*. If carefully observed, *S* is negatively correlated to total extractable P, $[P]_T$, exchangeable Ca, Mg, K and sum of bases in most study sites, with the exception being a small group of plots, all occurring on very weathered soils (mostly Ferralsols, Arenosols and some Acrisols) which show different behaviour from the remaining areas. The different behaviour of these sites on highly weathered soils may be related to the predominance of slow growth strategy which selects fewer and bigger trees. We also observe that study sites with very high *S* were often associated to extremely poor soils such as Podzols. Also of note is that all forests in our dataset had *S* above 450 trees ha^{-1} which is the minimum *S* associated to Amazonian forests in our dataset. Only one study site had *S* below that level (389 trees ha^{-1}) with this associated to a bamboo dominated forest over an extremely restrictive soil with regard to physical limitation.

No soil physical parameter result significantly correlated with *S* (Table S3). Despite of no significant correlation there are however interesting non-linear relationships with soil depth, soil structure and anoxic scores (Fig. S9). *S* seems to increase along with constraining soil conditions (i.e. increasing scores), but once soils become physically very restrictive (scores 3 and 4) then *S* decreases. Similar relationship also occurs for Π_2.

The strongest relationships for *S* occur with climate (Table S3, Fig. S10). There is a strong linear relationship between *S* and precipitation (ρ 0.332 p=0.002) and dry season precipitation (ρ 0.389 p<0.001). The number of months in which available water content was modelled to reached values below 20% was also negatively correlated to *S* (ρ -0.244 p=0.034), all these suggesting that water availability may be critical to determine the maximum number of trees that may occur on a given area.

The best OLS model fit for *S* included readily available P ($[P]_a$), $[Mg]_E$, $[K]_E$, T_A and P_D as predictors (Table. S2). The regression model resulted highly significant (p<0.001), explaining about to 42% of variation in *S*. All predictors turned out significant (p=0.024, p=0.002, p<0.001, p=0.004 and p<0.001, respectively). Only one valid alternative model could be selected within the ΔAIC<2 interval, with this having an ΔAIC of 0.970. This model included $[P]_a$, $[Ca]_E$, $[K]_E$, T_A and P_A as predictors, all of which attained similar levels of significance.

Model selection including the SEVM-1 filters resulted in $[Ca]_E$, $[Mg]_E$ and P_D as predictors to *S*, along with eight eigenvector filters. All variable resulted significant (p=0.005, p=0.011 and p<0.001,

respectively). However, the inclusion of SEVM-1 filters into the OLS model resulted in [P]a and TA as not significant, while $[Mg]_E$, $[K]_E$ retained their significance and P_D resumed as p=0.090.

The inclusion of SEVM-2 filters into model selection resulted in Π_1, ($[P]_a$), $[K]_E$ and P_D with all but Π_1 being significant in that model (p=0.790, p=0.011, p=0.011 and p=0.023, respectively). The addition of SEVM-2 filters into the OLS regression also resulted in P_D as not significant (p=0.214) but with all other variables resuming significant. No SEVM-3 filter has been selected for *S*.

Our results suggest that both precipitation and soil fertility may drive variations in number of stems.

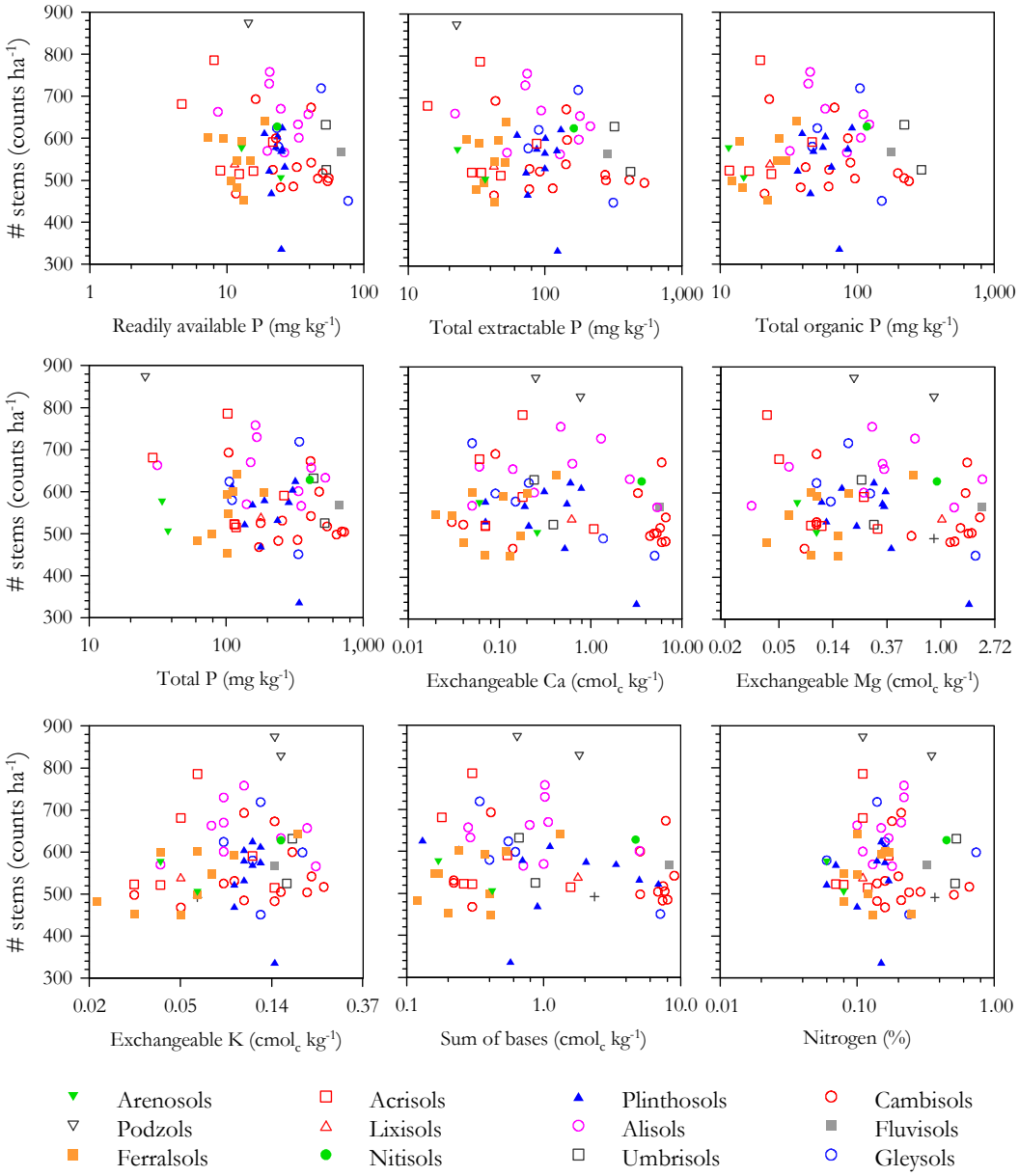

Figure S8. Relationships between stem density and different soil fertility parameters

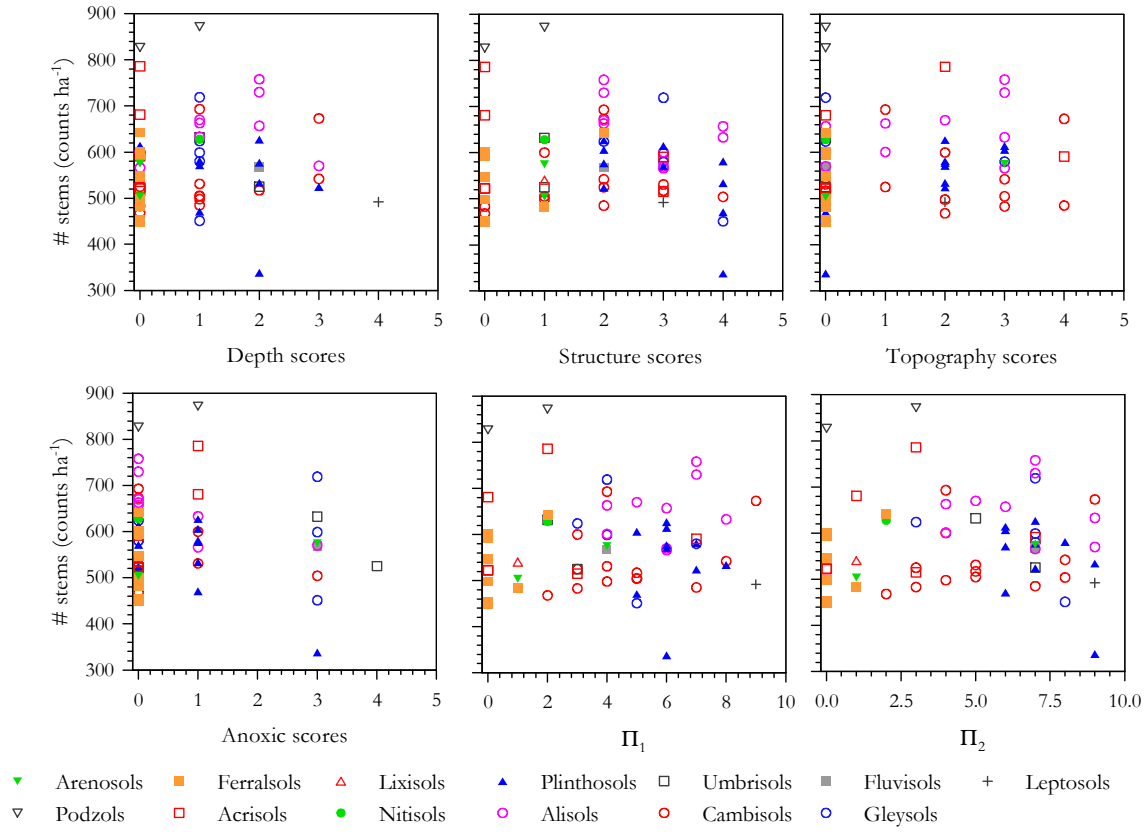

Figure S9. Relationships between stem density and soil physical properties

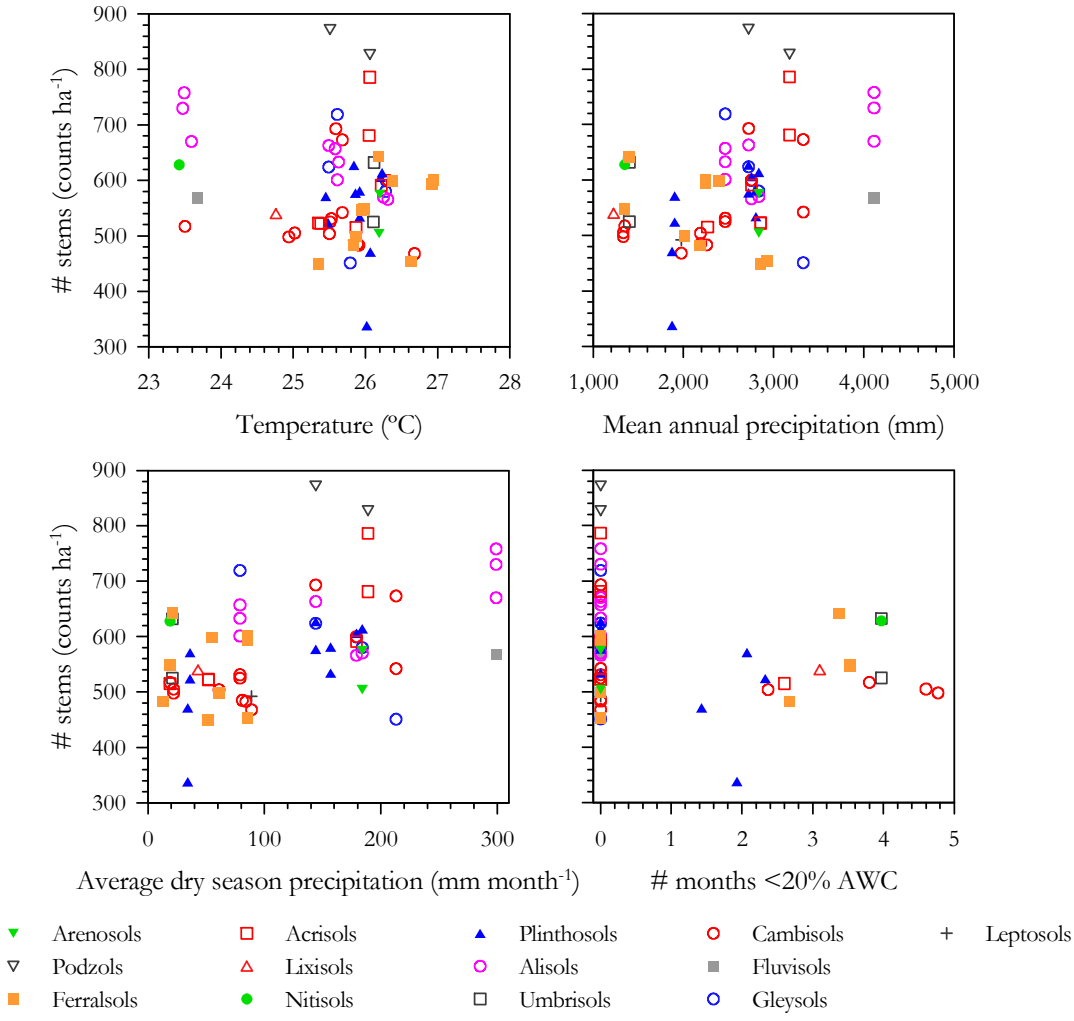

Figure S10. Relationships between stem density and climatic factors

Table S3 Spearman correlations between the number of stems per hectare and different predictors, adjusted for spatial autocorrelation by Dutillieul's method (Dutillieul, 1993).

Variable	Spearman rho	p value	p value adj	df adj
Sand	-0.008	0.511	0.581	41.03
Clay	0.123	0.809	0.829	46.25
Silt	-0.014	0.411	0.467	45.54
Soil depth	0.075	0.529	0.523	59.81
Soil structure	0.001	0.613	0.663	43.15
Topography	0.162	0.162	0.261	39.85
Anoxic	-0.007	0.544	0.565	52.20
Π_1	0.108	0.577	0.639	41.07
Π_2	0.101	0.795	0.832	38.74
Temperature	-0.104	0.075	0.098	50.03
Precipitation	**0.332**	**0.002**	**0.045**	**23.91**
Dry s. min. precip	**0.389**	**<0.001**	**0.027**	**22.38**
# month <20% AWC	**-0.244**	**0.034**	**0.169**	**24.70**
pH	-0.053	0.214	0.301	40.28
P_{TE}	-0.078	0.139	0.260	33.75
$P_{readily\ available}$	-0.127	0.187	0.308	34.82
P_{inorg}	-0.153	0.079	0.141	40.96
P_{org}	-0.043	0.258	0.407	31.36
P_{total}	-0.124	0.139	0.214	40.99
Nitrogen	0.021	0.590	0.667	37.28
Carbon	0.062	0.982	0.985	41.41
C:N	0.011	0.069	0.113	44.25
Ca	**-0.119**	**0.038**	**0.096**	**37.39**
Mg	**-0.112**	**0.069**	**0.109**	**45.03**
K	0.144	0.580	0.632	43.54
Al	0.311	0.248	0.350	38.14
Sum of bases	**-0.085**	**0.043**	**0.100**	**38.48**
ECEC	-0.007	0.128	0.185	44.13

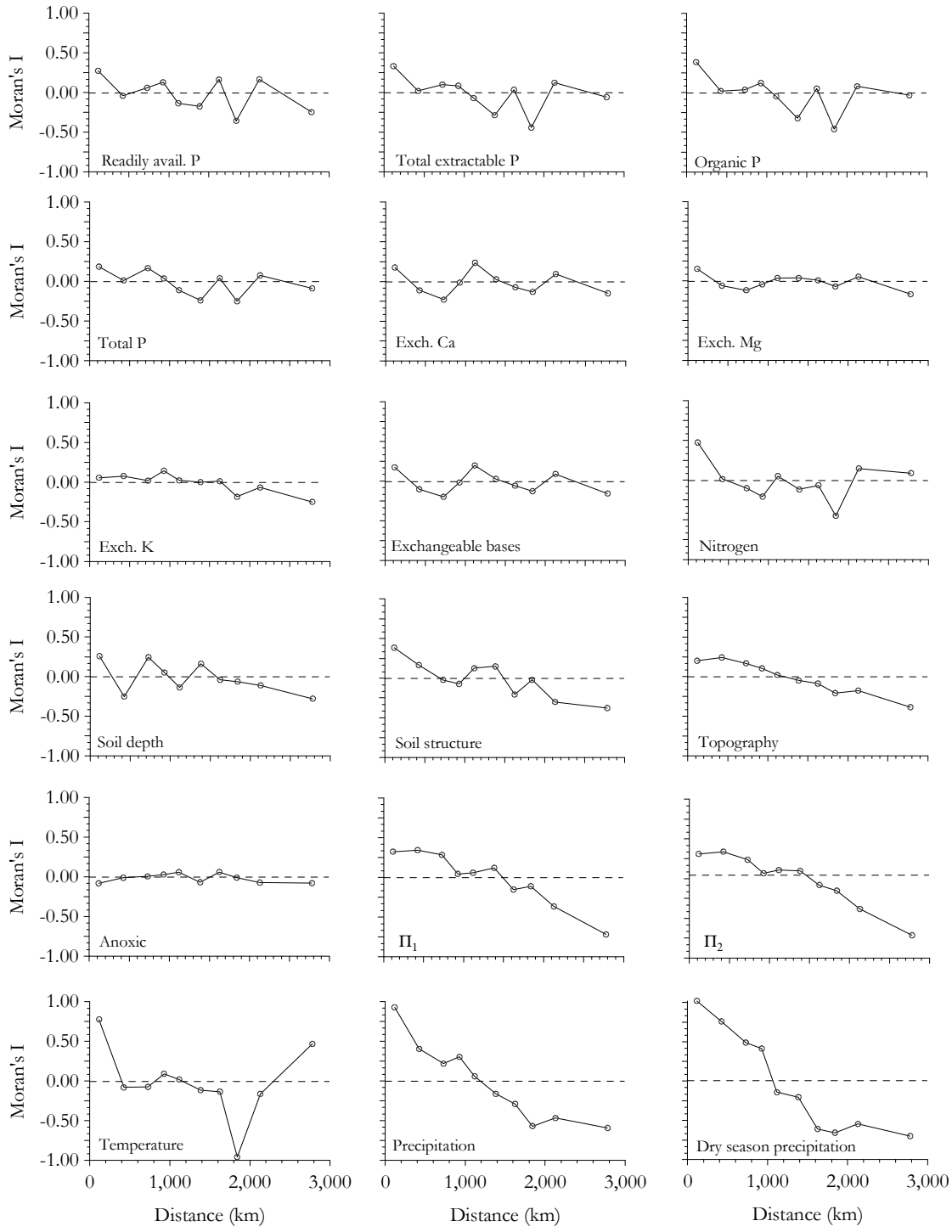

Fig. S11. Spatial autocorrelation levels (Moran's I) for the most important environmental predictors in this study.