Supplementary Materials

S1. Preliminary Analysis

We conducted a series of preliminary analyses before developing a predictive model of species richness that only used 1.0-ha scale data. We parameterized the preliminary models with lidar and field data from four spatial scales: 0.01-ha, 0.04-ha, 0.25-ha, and 1.0-ha (Figure S1). The preliminary analysis involved three main steps: (1) fitting species richness models to the five possible predictor variables (MCH, SDCH, mean slope, mean elevation, mean curvature); (2) using generalized least squares models with or without exponential correlation structure and nugget effect to model spatially autocorrelated residuals when present [1–4]; and (3) testing for spatially autocorrelated residuals using Moran's I [5].

We began with a set of four candidate models: null (intercept-only), topography (mean elevation, mean slope, and mean curvature), structure (mean canopy height, standard deviation of canopy height), and a "full" model with the terms from both the topography and structure models. We fit the models using both non-spatial and spatial GLS at each spatial scale separately. Spatial GLS extends OLS by including a covariance matrix for covariance of residuals between locations. In all but simulation situations [2], the covariance structure is unknown, so it is approximated by iteratively fitting the coefficients to the predictors and a variogram model to the residuals. The non-spatial GLS models were equivalent to OLS models. Briefly, the formula for GLS is:

$$Y = X\beta + \eta \quad (1)$$

where $\sim N(0, \sigma^2\Sigma)$, σ^2 is the variance, and Σ is the covariance matrix, and N is an i.i.d. normally distributed random variate with mean zero and variance of $\sigma^2\Sigma$. For comparison, the formula for OLS is:

$$Y = X\beta + \varepsilon \quad (2)$$

where $\varepsilon \sim N(0, \sigma^2 \mathrm{I})$, σ^2 is the variance, and I is the identity matrix. The non-spatial GLS model effectively simplified to an OLS model. It can be mathematically proven [3,4] that OLS provides unbiased estimates of GLS coefficients, but confidence intervals will be incorrect for OLS models when there is residual spatial autocorrelation. Our results across spatial scales follow from this fact, there was little difference between non-spatial and spatial GLS coefficient estimates, though spatial GLS models in general had wider 95% confidence intervals for coefficient estimates (Table S1). We detected spatial autocorrelation in non-spatial full model residuals at all spatial scales except 1.0-ha (Table S2). We evaluated the residual spatial autocorrelation for both non-spatial and spatial GLS models using Moran's I statistic. We observed that the spatial models were sufficient to control for residual spatial autocorrelation in all models at all spatial scales except 0.01-ha, where all but the full model still had residual autocorrelation after fitting an exponential correlation structure. We performed model averaging [6,7] based on small sample size Akaike's Information Criteria (AICc) to derive model average coefficients, standard errors, and confidence limits for the coefficient estimates (Table S3).

Figure S1. Species Richness in the Barro Colorado Island 50-ha Forest Dynamics Plot, Panama: (**a**) 0.01-ha; (**b**) 0.04-ha; (**c**) 0.25-ha; and (**d**) 1.0-ha.

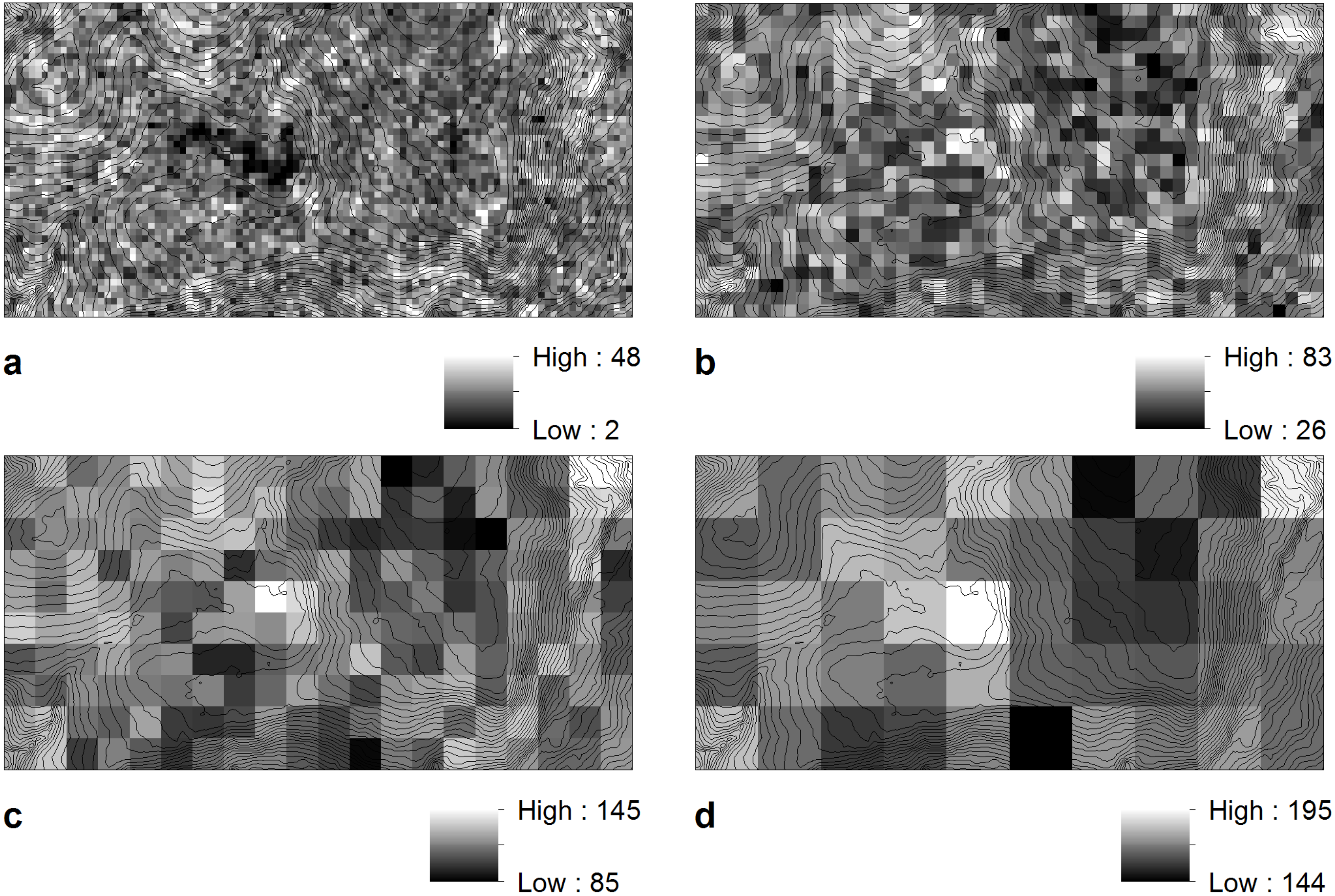

Table S1. Coefficient Estimates for species richness models. Significant terms are shown in black and non-significant terms are shown in grey.

	Parameter	Non-Spatial				Spatial			
		Estimate	SE	95% CI		Estimate	SE	95% CI	
				Upper	Lower			Upper	Lower
0.01-ha	SDCH	0.18	0.03	0.24	0.11	0.07	0.03	0.13	0.007
	MCH	−0.08	0.01	−0.06	−0.10	−0.05	0.01	−0.02	−0.08
	Elevation	−0.038	0.01	−0.02	−0.06	−0.04	0.02	0.004	−0.08
	Slope	0.22	0.02	0.26	0.18	0.13	0.04	0.20	0.06
	Curvature	0.28	0.07	0.41	0.15	0.20	0.06	0.31	0.08
	Intercept	27.4	1.7	30.8	24.1	28.0	3.7	35.2	20.7
0.04-ha	SDCH	0.600	0.11	0.82	0.38	0.40	0.11	0.61	0.19
	MCH	−0.302	0.04	−0.23	−0.38	−0.25	0.04	−0.17	−0.33
	Elevation	−0.0051	0.03	0.01	−0.11	−0.054	0.06	0.07	−0.17
	Slope	0.66	0.07	0.80	0.52	0.503	0.10	0.70	0.31
	Curvature	0.847	0.39	1.62	0.07	0.76	0.35	1.46	0.07
	Intercept	58.1	5.4	68.7	44.4	59.3	9.9	78.8	39.9
0.25-ha	SDCH	2.32	0.466	3.24	1.41	1.68	0.45	2.56	0.79
	MCH	−0.941	0.154	−0.64	−1.24	−0.95	0.16	−0.64	−1.25
	Elevation	−0.097	0.094	0.09	−0.28	−0.065	0.157	0.24	−0.37
	Slope	1.022	0.232	1.48	0.58	0.991	0.278	1.53	0.45
	Curvature	−0.785	2.74	4.58	−6.15	1.34	2.56	6.36	−3.68
	Intercept	127.7	16.6	160.3	95.1	127.8	25.8	178.3	77.3
1.0-ha	SDCH	2.77	1.15	5.02	0.53	2.75	1.14	4.99	0.51
	MCH	−1.072	0.408	−0.27	−1.87	−1.09	0.412	−0.29	−1.90
	Elevation	−0.313	0.519	0.05	−0.68	−0.277	0.203	0.12	−0.67
	Slope	0.418	0.519	1.43	−0.60	0.566	0.534	1.61	−0.48
	Curvature	−32.7	11.16	−10.8	−54.6	−30.9	11.3	−8.78	−53.0
	Intercept	218.6	33.9	285.0	152.3	212.0	36.7	283.8	140.1

Table S2. Results of Moran's I test for residual spatial autocorrelation in non-spatial and spatial GLS models of species richness parameterized with BCI 50 ha plot census data.

Scale	Model	Variogram		Moran's I							
				Non-Spatial				Spatial			
		Range	Nugget	Observed	Expected	SD	*p-value*	Observed	Expected	SD	*p-value*
0.01-ha	Null	12.2	0.231	0.0242	−0.0002	0.0004	<0.0001	0.0009	−0.0002	0.0004	0.0038
	Topography	11.2	0.207	0.0177	−0.0002	0.0004	<0.0001	0.0004	−0.0002	0.0004	0.0860
	Structure	12.1	0.244	0.0237	−0.0002	0.0004	<0.0001	0.0010	−0.0002	0.0004	0.0014
	Full	10.8	0.212	0.0158	−0.0002	0.0004	<0.0001	0.0005	−0.0002	0.0004	0.0663
0.04-ha	Null	23.6	0.193	0.0408	−0.0008	0.0013	<0.0001	0.0011	−0.0008	0.0013	0.1443
	Topography	17.7	0.034	0.0342	−0.0008	0.0013	<0.0001	0.0018	−0.0008	0.0013	0.0494
	Structure	31.0	0.372	0.0400	−0.0008	0.0013	<0.0001	0.0005	−0.0008	0.0013	0.3286
	Full	19.9	0.221	0.0281	−0.0008	0.0013	<0.0001	0.0010	−0.0008	0.0013	0.1632
0.25-ha	Null	62.5	0.180	0.0498	−0.0050	0.0063	<0.0001	−0.0028	−0.0050	0.0063	0.7255
	Topography	76.8	0.359	0.0496	−0.0050	0.0063	<0.0001	−0.0040	−0.0050	0.0063	0.8732
	Structure	75.9	0.295	0.0453	−0.0050	0.0063	<0.0001	−0.0033	−0.0050	0.0063	0.7827
	Full	81.6	0.460	0.0382	−0.0050	0.0063	<0.0001	−0.0045	−0.0050	0.0063	0.9315
1.0-ha	Null	80.5	0.248	0.0212	−0.0204	0.0189	0.0276	−0.0209	−0.0204	0.0189	0.9791
	Topography	5.6	0.121	−0.0048	−0.0204	0.0189	0.4114	−0.0048	−0.0204	0.0189	0.4114
	Structure	71.4	0.128	−0.0242	−0.0204	0.0189	0.0190	−0.0199	−0.0204	0.0189	0.9783
	Full	83.8	0.756	0.0067	−0.0204	0.0189	0.1551	−0.0077	−0.0204	0.0189	0.5051

Table S3. Multimodel Inference Table for species richness GLS models. AIC_c is small sample size Akaike's Information Criterion. Models are ranked based on AIC_c. Δ_i is the difference in AIC_c from the rank 1 model and w_i is the model probability given the data and set of candidate models.

	Non-Spatial						Spatial					
Rank	**Model**	**No. Parameters**	**Log Likelihood**	**AIC_c**	**Δ_i**	**w_i**	**Model**	**No. Parameters**	**Log Likelihood**	**AIC_c**	**Δ_i**	**w_i**
1	Full	7	−15,396.1	30,806.2	0.00	1.00	Full	9	−14,930.3	29,878.6	0.00	1.00
2	Topography	5	−15,456.7	30,923.3	117.1	0.00	Topography	7	−14,942.7	29,899.4	20.8	0.00
3	Structure	4	−15,488.0	30,984.1	177.9	0.00	Structure	6	−14,945.3	29,902.6	24.0	0.00
4	Null	2	−15,540.3	31,084.6	278.4	0.00	Null	4	−14,956.8	29,921.5	42.9	0.00
1	Full	7	−4,377.3	8,768.6	0.00	1.00	Full	9	−4,281.4	8,581.0	0.00	1.00
2	Topography	5	−4,430.2	8,870.4	101.7	0.00	Structure	7	−4,295.7	8,603.4	22.4	0.00
3	Structure	4	−4,435.6	8,879.2	110.5	0.00	Topography	6	−4,311.9	8,637.9	56.9	0.00
4	Null	2	−4,477.6	8,959.2	190.6	0.00	Null	4	−4,326.3	8,660.7	79.7	0.00
1	Full	7	−716.6	1447.8	0.00	1.00	Full	9	−703.1	1,425.2	0.0	0.98
2	Structure	4	−731.3	1470.9	23.1	0.00	Structure	7	−710.3	1,433.0	7.8	0.02
3	Topography	5	−742.9	1496.0	48.2	0.00	Topography	6	−725.4	1,465.3	40.11	0.00
4	Null	2	−753.0	1510.0	62.2	0.00	Null	4	−730.5	1,469.2	43.98	0.00
1	Full	7	−170.2	357.1	0.00	0.97	Full	9	−170.0	362.5	0.00	0.96
2	Topography	5	−176.6	364.6	7.5	0.02	Topography	7	−176.6	369.9	7.4	0.02
3	Structure	4	−180.3	369.5	12.4	0.00	Structure	6	−178.2	370.3	7.9	0.02
4	Null	2	−186.4	377.1	20.0	0.00	Null	4	−184.2	377.3	14.9	0.00

Table S4. The proportion of explained variance in the non-spatial and spatial GLS models of species richness parameterized with BCI 50 ha plot census data.

Scale	**R^2**	
	Non-Spatial	**Spatial**
0.01-ha	0.06	0.05
0.04-ha	0.15	0.14
0.25-ha	0.31	0.28
1.0-ha	0.49	0.49

We evaluated the proportion of explained variance in the full model as an upper boundary for the proportion of explained variance given the variables and type of model (spatial or non-spatial), at each spatial scale, by the proportion of predictive variance over the total variance. This is equivalent to the coefficient of determination (R^2) in OLS, but for GLS there is no analytical formula for determining significance of R^2. We evaluated the proportion of explained variance at each spatial scale to identify the scale at which the association of species richness to the lidar variables was strongest. Based on the preliminary analysis it was clear that species richness is associated with canopy height and terrain structure variables at all spatial scales, but most strongly at the coarsest scale, 1.0-ha (Table S4). The proportion of explained variance increased from fine (0.01-ha) to coarse (1.0-ha) spatial scale.

S2. Leave-One-Out Cross-Validation

The algorithm for leave-one-out cross-validation [8] is: (1) divide the data into training and testing set. Let N be the number of unique observations. The training set is N-1 and the testing set is 1 datum; (2) Fit the model to the training set and use the test datum to predict the response variable; (3) evaluate the difference between the observed and predicted response variable; (4) repeat for all N unique test datum.

References

1. Cressie, N. *Statistics for Spatial Data*; Wiley: New York, NY, USA, 1991; pp. 1–900.
2. Beale, C.M., J.J. Lennon, J.M. Yearsley, M.J. Brewer, D.A. Elston regression analysis of spatial data. *Ecol. Lett.* **2010**, *13*, 246–264.
3. Draper, N.R.; Smith, H. *Applied Regression Analysis*; 3rd ed.; Wiley: New York, NY, USA, 1998; pp. 1–736.
4. Baltagi, B.H. *Econometrics*; 5th ed.; Springer: New York, NY, USA, 2011; pp. 1–425.
5. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23.
6. Burnham, K.P.; Anderson, D.R.; Huyvaert, K.P. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **2011**, *65*, 23–35.
7. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; 2nd ed.; Springer: New York, NY, USA, 2002; pp. 1–488.
8. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Statistics Surveys* **2010**, *4*, 40–79.