Strong spatial and temporal limitations in seed arrival as complementary mechanisms for species coexistence in a tropical Atlantic coastal forest

Supplementary Information

Leticia B. Zimback[1,2], Paulo I. Prado[2], Marcelo P. Pansonato[2], Geraldo A. D. C. Franco[3], and Adriana M. Z. Martini[2,4]

[1]Secretaria do Verde e Meio Ambiente da Prefeitura da Cidade de São Paulo, Brazil
[2]Instituto de Biociências, Universidade de São Paulo, Brazil
[3]Instituto Florestal do Estado de São Paulo, Brazil
[4]Corresponding author, `amzmartini@ib.usp.br`

Submitted to *Plant Ecology*

This appendix provides additional information on the statistical methods used in the paper and also some supplementary analyses. All analyses were performed in the R environment [12]. R codes and the dataset are available at `https://github.com/piklprado/seedrain`.

Linear models for spatial and temporal seed limitation

In this section we report details of the statistical models used to describe spatial seed limitation (*SSL*, equation 1 in Methods in the main paper) and temporal seed limitation (*TSL*, equation 2). We also provide the coefficients of the selected models (ΔAICc $\leq$ 2), AICc values for each model and the coefficients for the average model, obtained from the selected models.

Generalized linear mixed models (glmm) were used to evaluate which traits were more related to the spatial and/or temporal seed limitation of the species sampled. As the response variables (SSL and TSL) are the number of unoccupied traps or months out of a total, a binomial glmm with the logit link function was used [5, 7]. Through the inclusion of random factors, a mixed model

approach allows accounting for hierarchical relationships and non-independent factors adequately [5, 13]. In these models, species identity was included as a random factor, as there were three values for each species (i.e., one for each year), and it was expected that temporal or spatial limitation of the same species was non-independent among years.

Fixed factors used in the linear models were the following:

Log seed mass : logarithm dry seed mass average of each plant species. Dry seed mass average were calculated from at least 10 seeds randomly picked among those collected in the traps (for most abundant species, 50 seeds were used);

Adult maximum height : maximum tree height recorded in the study site [11].

Frequency of adults : proportion of the 40 plots in which a tree species was observed in the study site inventory [11].

Frequency of adult trees was included to discount the effect of the spatial distribution of adult individuals on seed limitation. As seed shadow is strongly concentrated below or nearby adult trees [8, 9], the presence of a species in a given seed trap is more probable where an adult individual is present in the plot. Hence, for spatial seed limitation it would be expected that a species with a wide spatial distribution of adult individuals would show low spatial seed limitation. The relationship between frequency of adult trees and temporal seed limitation is not as direct. However, species with a broad spatial distribution might be subjected to more heterogeneous environmental conditions. Consequently, individuals under distinct conditions (for example, canopy gaps/closed canopy, flooding/non-flooding) could promote intraspecific asynchrony among adult individuals of a species and extend its reproductive period. Hence, a broad spatial distribution could be related to low temporal seed limitation. Seeds were also previously classified into two dispersal syndrome categories, zoochorous and non-zoochorous. However, only four species (from 31 at total) had non-zoochorous seeds. As this strongly unbalanced distribution precludes a robust analysis of its effects on seed limitation, this trait was not used in the analyses.

For each response variable — spatial seed limitation (SSL) or temporal seed limitation (TSL) — a full glmm model was fit with the glmer function from the lme4 package [3]. In the full model, species identity was included as a random factor, and all other variables and their two-way interactions were included as fixed factors. Continuous fixed factors were standardized by computing z-score values to ease model convergence [13] and to allow for a direct comparison of their effects. By using the *dredge* function from the MuMIn package [2], all possible combinations of fixed factors and two-way interactions were fit and compared. Model selection was performed using Akaike information criteria (with correction for small samples) that were calculated with the *AICc* function from the bbmle package [4]. All models with AICc $\leq$ 2.0 were considered

equally plausible [6]. When multiple models were selected, the average model was calculated [6] using the *model.avg* function from the MuMIn package [2] to estimate the effect of each fixed factor and confidence intervals. Predicted values of the averaged model were estimated as the average of the fitted values of each selected model, weighted by their Akaike evidence weights. Pseudo$-R^2$ [10] values were calculated for each selected model using the *r.squaredGLMM* function from the MuMIn package.

Additionally, generalized linear models (glm) were fitted considering the total pooled for the 36 months, for both SSL and TSL. For these analyses, a mixed model was not required because they did not have repeated values for each species. The procedures of model expansion using the function *dredge* and model selection were performed in the same way as described above for the glmm models, but applying quasi-binomial fitting and quasi-AIC for the model selection were needed to correct overdispersion of the binomial fits [13].

Some species/morphotypes were excluded from the correlation analysis and from the trait relationships analysis: i) allochthonous seeds (i.e., seeds of species not recorded as adult trees within the forest fragment); ii) morphotypes identified only to the genus level, and with two or more species of this genus among the adults of the forest site; and iii) species with less than 5 seeds sampled along the three years, which was stated as a minimum to obtain reliable estimates of seed limitation.

Tables S-1 to S-4 report the coefficients of the selected models, AICc values for each model and the coefficients for the average model, obtained from the selected models. All the fixed-effect variables were standardized as z-values, to ease convergence of the mixed-effect models, and also to allow comparison of the coefficients for each effect. We thus refer the coefficients as "standardized coefficients".

Figures S-1 and S-2 show the observed values of pooled SSL and TSL respectively, and the predicted values by the fitted generalized linear models.

Circular statistics of temporal distribution of seeds

The mean number of seed species in each month for the three years was used to calculate a mean distribution vector (r), represented by the red line in Fig. S-3. Small vectors, like the one seen in Fig. S-3, indicate low temporal synchrony and high uniformity among the months. The calculated value of the mean distribution vectors was $r = 0.0475$, and the Rayleigh Test of Uniformity (Z) returned a p-value of $p = 0.709$, and thus the null hypothesis of uniform distribution over months can not be rejected. This figure, the r calculation, and the Rayleigh test were performed with the circular R package [1].

Table S-1: Mixed-effect models for the spatial seed limation. For each model is shown standardized coefficients and standard error (brackets). Also shown the estimated standard deviation for the random effects (SD), the AICc value for each selected model, and the coefficients of the average model.

	8	16	7	24	32	6	Average
Adult Frequency	−0.410	−0.303		−0.445	−0.334	−0.618	−0.341
	(0.226)	(0.231)		(0.224)	(0.227)	(0.218)	(0.270)
Adult height	−0.478	−0.542	−0.659	−0.467	−0.536		−0.473
	(0.231)	(0.227)	(0.217)	(0.226)	(0.221)		(0.277)
Log seed mass	1.018	1.024	0.975	1.003	1.009	1.029	1.010
	(0.205)	(0.197)	(0.214)	(0.201)	(0.192)	(0.220)	(0.205)
Frequency:Height		−0.320			−0.335		−0.106
		(0.234)			(0.228)		(0.202)
Frequency:Mass				0.257	0.268		0.074
				(0.224)	(0.214)		(0.166)
SD (Intercept)	1.022	0.979	1.085	1.002	0.952	1.113	
SD (Observations)	1.000	1.000	1.000	1.000	1.000	1.000	
AICc	489.753	490.276	490.645	490.769	491.099	491.440	

Table S-2: Fixed effects models for the spatial seed limitation pooled over years (see text for details). For each model is shown standardized coefficients and standard error (brackets). Also shown the QAICc value for each selected model, and the coefficients of the average model.

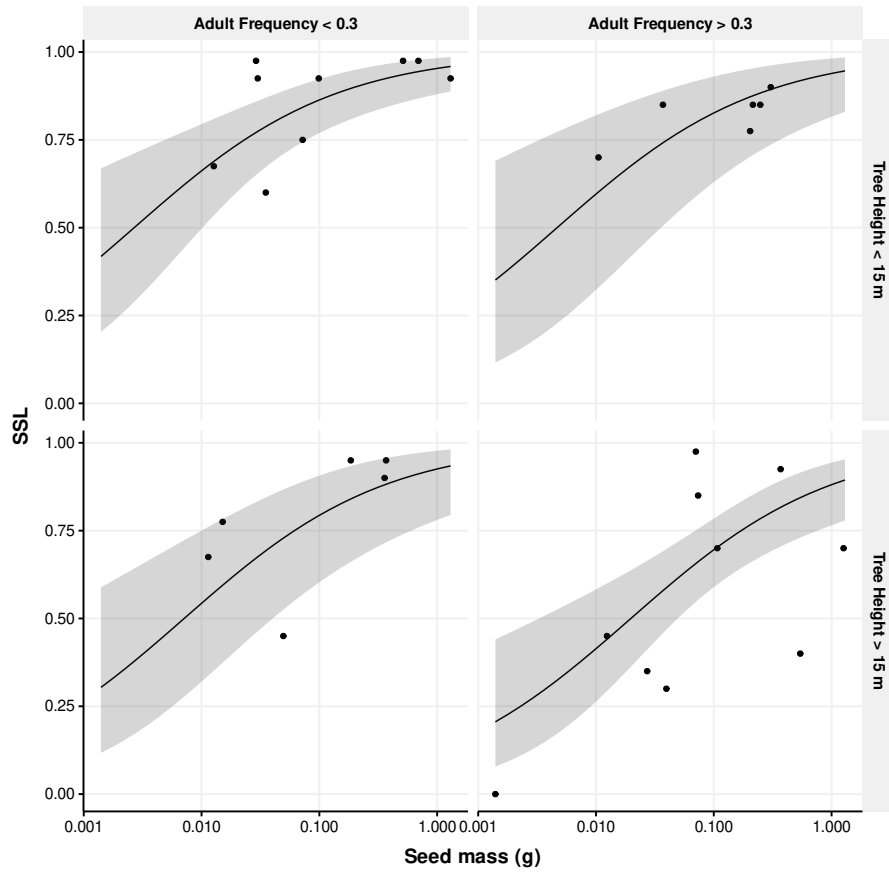|	8	16	7	6	Average
Adult Frequency	−0.363	−0.164		−0.538	−0.257
	(0.203)	(0.241)		(0.198)	(0.269)
Adult height	−0.432	−0.528	−0.597		−0.413
	(0.227)	(0.238)	(0.210)		(0.293)
Log seed mass	0.836	0.851	0.771	0.849	0.826
	(0.206)	(0.198)	(0.199)	(0.224)	(0.208)
Frequency:Height		−0.395			−0.100
		(0.237)			(0.209)
QAICc	56.540	56.920	56.965	57.439	

Figure S-1: Relationship among spatial seed limitation pooled over years (SSL), seed mass, tree height, and frequency of adults, as predicted by the glm average model (Table S-4). For species with frequency lower than median value (0.3) the relationship is showed in graphs in the left column, and for those with greater frequency, in the right column. For species with tree height lower than median value (15 m) the relationship is showed in upper graphs, and for those with greater height in bottom graphs. Regression lines are predicted values by the average model using the midpoint of the height and frequency class in each panel. Gray shadows are 95% prediction interval of the average model. Points are observed values of SSL (pooled over the 3 years) and seed mass for each species. Note the log scale for seed mass.

Table S-3: Mixed-effect models for the temporal seed limation. For each model is shown standardized coefficients and standard error (brackets). Also shown the estimated standard deviation for the random effects (SD), the AICc value for each selected model, and the coefficients of the average model.

	7	6	5	8	Average
Adult Frequency		−0.295		−0.174	−0.097
		(0.195)		(0.211)	(0.179)
Adult height	−0.355			−0.278	−0.180
	(0.194)			(0.213)	(0.223)
Log seed mass	0.971	1.000	0.963	0.992	0.979
	(0.197)	(0.201)	(0.206)	(0.197)	(0.200)
SD (Intercept)	0.919	0.936	0.981	0.905	
SD (Observations)	1.000	1.000	1.000	1.000	
AICc	356.252	357.240	357.249	357.819	

Table S-4: Fixed effects models for the temporal seed limitation pooled over years (see text for details). For each model is shown standardized coefficients and standard error (brackets). Also shown the QAICc value for each selected model, and the coefficients of the average model.

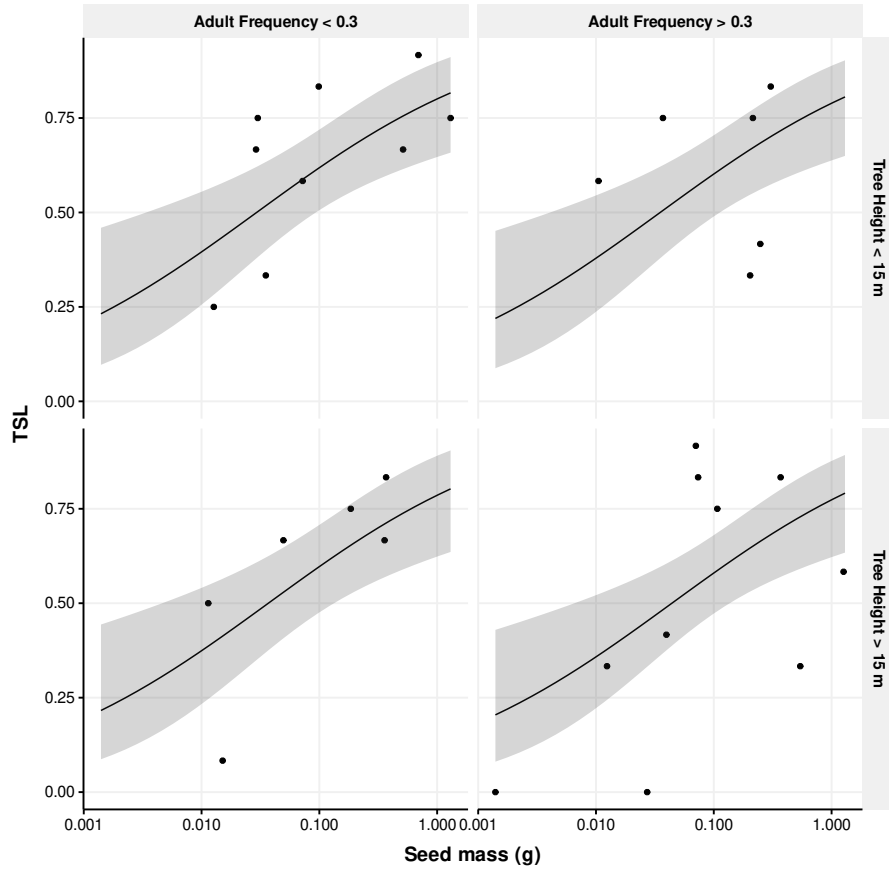|	5	7	6	Average
Adult Frequency			−0.198	−0.045
			(0.182)	(0.120)
Adult height		−0.224		−0.059
		(0.185)		(0.137)
Log seed mass	0.628	0.639	0.661	0.638
	(0.193)	(0.194)	(0.198)	(0.194)
QAICc	62.212	63.548	63.812	

Figure S-2: Relationship among temporal seed limitation pooled over years (TSL), seed mass, tree height, and frequency of adults, as predicted by the glm average model (Table S-3). For species with frequency lower than median value (0.3) the relationship is showed in graphs in the left column, and for those with greater frequency, in the right column. For species with tree height lower than median value (15 m) the relationship is showed in upper graphs, and for those with greater height in bottom graphs. Regression lines are predicted values by the average model using the midpoint of the height and frequency class in each panel. Gray shadows are 95% prediction interval of the average model. Points are observed values of SSL (pooled over the 3 years) and seed mass for each species. Note the log scale for seed mass.
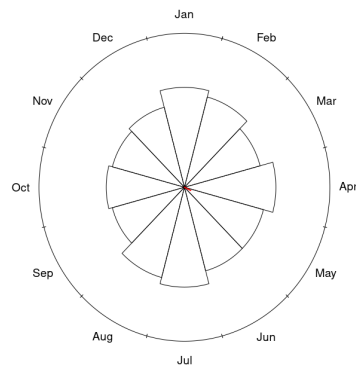
Figure S-3: Circular distribution of the mean number of seed species in each month. The lenght of each sector (month) is the average of the number of seed species recorded each month for the tree years. Maximum lenght is 16,67 for July.

References

[1] C. Agostinelli and U. Lund. *R package* `circular`: *Circular Statistics (version 0.4-94)*. CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA, 2022.

[2] Kamil Bartoń. *MuMIn: Multi-Model Inference*, 2020. R package version 1.43.17.

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[4] Ben Bolker and R Development Core Team. *bbmle: Tools for General Maximum Likelihood Estimation*, 2021. R package version 1.0.24.

[5] Benjamin M Bolker, Mollie E Brooks, Connie J Clark, Shane W Geange, John R Poulsen, M Henry H Stevens, and Jada-Simone S White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24(3):127–135, 2009.

[6] K P Burnham and D R Anderson. *Model Selection and Multimodel Inference - A Practical-Theoretic Approach*. Springer-Verlag, Amsterdam, 2002.

[7] Annette J Dobson and Adrian G Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.

[8] Habacuc Flores-Moreno, Fiona J Thomson, David I Warton, and Angela T Moles. Are introduced species better dispersers than native species? a global comparative study of seed dispersal distance. *PloS one*, 8(6):e68541, 2013.

[9] Eugenio Larios and David Lawrence Venable. Selection for seed size: the unexpected effects of water availability and density. *Functional Ecology*, 32(9):2216–2224, 2018.

[10] Shinichi Nakagawa, Paul CD Johnson, and Holger Schielzeth. The coefficient of determination r 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134):20170213, 2017.

[11] Marcelo Petratti Pansonato, Renato Augusto Ferreira de Lima, Alexandre Adalardo de Oliveira, Ricardo Bertoncello, and Adriana Maria Zanforlin Martini. Community structure and species composition of a periodically flooded restinga forest in caraguatatuba, são paulo, brazil. *Biota Neotropica*, 19, 2018.

[12] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2022.

[13] Alain Zuur, Elena N. Ieno, Neil Walker, Anatoly A. Saveiliev, and Graham M. Smith. *Mixed Effects Models and Extensions in Ecology with R.* Springer, New York, 2009. ISBN 978-0-387-87457-9.