A Parameter calibration

In the following we describe how we obtained the ecologically feasible values/ranges for the model parameters in (1).

- $A_{\max}$ (the adult tree carrying capacity) is estimated from Barro colorado Island (BCI) data reporting the total aboveground biomass densities of $28-30\,\mathrm{kg}\cdot\mathrm{m}^{-2}$ in[55]. We rounded this up to $30\,\mathrm{kg}\cdot\mathrm{m}^{-2}$.

- g_S (the seed growth rate) is computed as the ratio $g_S = \frac{g_S^{sp}\, g_S^{bc}}{A_{\max}}$, where g_S^{sp} is the average seed production and g_S^{bc} is the average seed weight. According to the data presented in[56], g_S^{sp} is around 0.15 seeds $\mathrm{y}^{-1}\cdot\mathrm{mm}^{-2}$, which corresponds to $15\cdot 10^4$ seeds $\mathrm{y}^{-1}\cdot\mathrm{m}^{-2}$. The mass of a single seed varies between $0.002-1\,\mathrm{g}$[57–59]. Therefore we obtain

$$g_S = \frac{15\cdot 10^4\cdot(0.002-1)\cdot 10^{-3}}{30}\,\mathrm{y}^{-1} = (0.01-5)\,\mathrm{y}^{-1}. \tag{7}$$

- k_S (seeds' turnover rate) is computed as an average total seed turnover rate of 1-6 y^{-1} based on a review of seed longevity as observed in tropical forests globally[60].

- g_N (the transition rate from seeds to seedlings) is given by the product of two factors $g_N = g_N^{tp}\cdot g_N^{bc}$, where:

  – The seed-to-seedling transition probability g_N^{tp} is calculated using BCI data as follows: $1-7\%$ of seeds in the seed bank germinates after gap formation[61]. Also, there is a difference of approximately 2-3 orders of magnitude between seed production and seed germination, which quickly follows production[62]. Therefore, we can conclude that 1% of seeds typically germinates. Assuming a maximum of 60% reduction of seedlings due to density-dependent effects, and that 1% germination is observed after these effects occurred, we obtain $g_N^{tp} = 0.01/0.4 = 0.025\,\mathrm{y}^{-1}$. (Note: We expect this value to be smaller than k_S.)

  – A biomass conversion factor g_N^{bc} due to the increase in biomass coinciding with the transition from seed to seedling . This conversion factor is given by the ratio between the biomass of a seedling and the biomass of a seed. The former is estimated to be 10 g on average (e.g.[63,64]); i.e. $g_N^{bc} = \frac{10}{(0.002-1)} = 10-5000$.

  Consequently, we have that

$$g_N = 0.025\cdot(10-5000)\,\mathrm{y}^{-1} = 0.25-125\,\mathrm{y}^{-1}. \tag{8}$$

- β and r_T (establishment sensitivity to toxicity parameters) correspond to parameters β and γ in[6], respectively. Therefore, $\beta = 10^{-5}$ and $r_T \in [0,68]$ (mimicking woody plants).

- k_N (death rate of the seedlings) is based on[65], according to which $2-74\%$ of the seedlings have died after two months. Therefore, $k_N \in [0.02, 0.74]$.

- r_P (the increased mortality of seedlings due to the inhibitor) is obtained by assuming that when I reaches its equilibrium values I^* (see Section B) the mortality increases by a factor $X * k_N$, with $0.3 < X < 0.9$[66,67]. Since the absolute maximum increased mortality observed in empirical studies, i.e. the reduction that happens when I reaches the theoretical maximum toxicity level I ($= A_{\max}$ with our parameter choice), is 90% (e.g.[68]), we get $0.2 < r_P < 0.7$.

- g_A (the transition rate from seedlings to adults is given by the relation $g_A = g_{SS}^{tp}\cdot g_{SA}^{tp}\cdot g_A^{bc}$, where

  – The seedling-to-sapling transition rate g_{SS}^{tp} is $0.0025-2\%$ according to[59].

  – The sapling-to-adult transition rate is estimated as $0.015-0.1\%$, based on[69–71].

  – The seedling-to-adult biomass conversion rate $g_{SA}^{tp}$is given by the ratio between the biomass of a sapling and the one of an seedling. In particular, the sapling biomass is assumed to span between 1 and 10 kg, with a median value of 5 kg[72–74].

  Therefore, we have

$$g_A = [(0.000025-0.02)\cdot(0.015-0.1)\cdot 5/0.01]\,\mathrm{y}^{-1} = 0.00019-1\,\mathrm{y}^{-1}. \tag{9}$$

- c_A (the growth rate in adults'biomass density) is calculated by imposing that the maximum growth rate (given by $\frac{c_A A_{\max}}{4}$, i.e. the logistic function $c_A A\left(1-\frac{A}{A_{\max}}\right)$ evaluated at the maximum $A=\frac{A_{\max}}{2}$) is equal to $2\,\mathrm{kg}\cdot\mathrm{m}^{-2}\cdot\mathrm{y}^{-1}$[75,76]. This gives $c_A=\frac{4}{15}\,\mathrm{y}^{-1}$, which can be rounded off to $0.25\,\mathrm{y}^{-1}$.
- k_A (the mortality rate of adult trees) is assumed to be proportional to the inverse of the average longevity of an adult, which is approximately $100\,\mathrm{y}$[77,78]. We then consider $k_A=0.01\,\mathrm{y}^{-1}$.
- c_T (the growth rate of inhibitor density due to adult tree density) is computed based on the ecologically reasonable assumption that the response time in the presence of an adult tree biomass is relatively rapid (order of half a year to year before soil equilibrates, see e.g.[79]). In average, we then get $c_T=0.7\,\mathrm{y}^{-1}$. We observe that $c_T>k_A$, since toxicity increases also when adults are living.
- k_I (the toxicity decay rate), analogously to c_T, is calculated based on the assumptions that the pathogens effect disappears quite quickly (around half a year to a year) after adult trees are removed. Therefore, we consider $k_I=0.7\,\mathrm{y}^{-1}$.
- d_S (the diffusion coefficient for seeds) was derived using empirically derived seed dispersal kernels[80]. Specifically, using we first simulated seed distribution patterns around a single adult tree based on a typical empirical seed dispersal kernel[80], and then approximating this distribution by a Gaussian. This Gaussian distribution could then be reproduced with a diffusion model, selecting the diffusion coefficient that provided the best fit to this Gaussian distribution. Through this procedure, we obtained best-fitting values within the range $d_S=3-4\,\mathrm{m}^2\,\mathrm{y}^{-1}$.
- d_I (the diffusion coefficient of toxicity) this spatial parameter is more challenging to parameterize than d_S, as direct observations are not available (in contrast to observed seed dispersal patterns). Hence, our general approach is to vary d_I, while keeping all other parameters fixed, to identify the range for which JC distributions emerge. To set the upper bound of the range to be considered, we relied on previous studies suggesting that negative density-dependent effects occur within a range of 30 m around adult trees[11,15,22]

B Steady-states

The steady-states of System (1) are given by the solutions to the following system

$$0=g_S\cdot A-k_S\cdot S, \tag{10a}$$

$$0=\frac{g_N\cdot S}{1+\beta\cdot e^{r_T\cdot I}}-\left(k_N+g_A\left(1-\frac{A}{A_{\max}}\right)+r_P\cdot I\right)\cdot N, \tag{10b}$$

$$0=(g_A\cdot N+c_A\cdot A)\cdot\left(1-\frac{A}{A_{\max}}\right)-k_A\cdot A, \tag{10c}$$

$$0=c_T\cdot A-k_I\cdot I. \tag{10d}$$

Because of the complexity deriving by the exponential term in the denominator of Equation (10b), this system has been carefully studied in its corresponding nondimensional form in[27], to which we refer for further details. For the purpose of the current paper, however, we merely recall that System (10) admits two solutions given by

$$E_0^*=(S_0^*,N_0^*,A_0^*,I_0^*)=(0,0,0,0)\,, \tag{11a}$$

$$E_1^*=(S_1^*,N_1^*,A_1^*,I_1^*)=\left(\frac{g_S\cdot A^*}{k_S},\frac{g_S f\left(\frac{A^*}{k_I}\right)}{k_S\left(k_N+\frac{r_P}{k_I}A^*+g_A\left(1-\frac{A^*}{A_{\max}}\right)\right)}A^*,A^*,\frac{c_T\cdot A^*}{k_I}\right), \tag{11b}$$

where A^* is the unique solution of

$$f\left(\frac{A}{k_I}\right)A=g(A)A,$$

with

$$f(X):=\frac{g_N}{1+\beta\cdot e^{r_T\cdot X}},$$

$$g(X):=\frac{k_S\left(c_A\left(\frac{X}{A_{\max}}-1\right)+k_A\right)\left(g_A\cdot X-A_{\max}\left(g_A+\frac{r_P\cdot X}{k_I}+k_N\right)\right)}{g_A\cdot g_S\,(X-A_{\max})}.$$

C Additional plots for S, A, and I

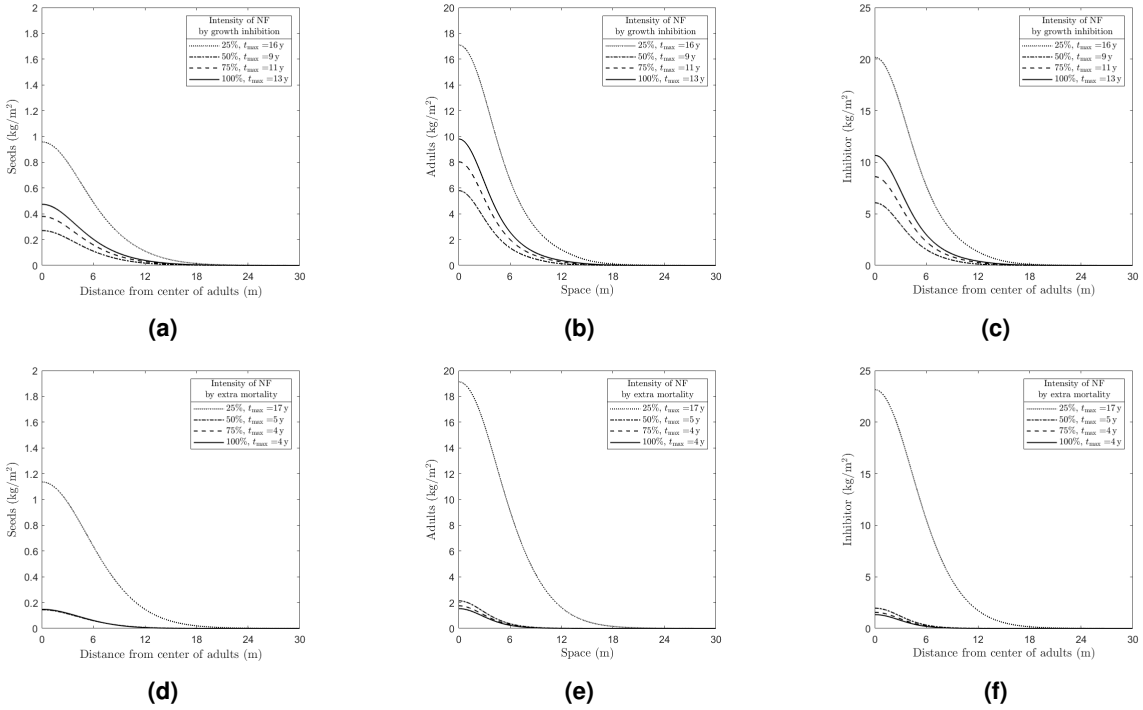

Figure 8. Numerical investigation of the NF influence on the emergence of Janzen-Connell distributions by means of exclusively growth inhibition (top row) and increased mortality (bottom row) for S ((a), (d)), A ((b), (e)), and I ((c), (f)). The plots show the variables' profiles at time $t = t_{\max}$ (where the maximum amplitude for N is reached, indicated in the legend) obtained by simulating System (1) for different values of the establishment sensitivity to autotoxicity parameter r_T, corresponding to different percentages of $r_T^{\text{ref}} = 68\,\text{m}^2\,\text{kg}^{-1}$ (top row), and different values of r_P representing the increased mortality induced by soil-borne pathogens, corresponding to different percentages of $r_P^{\text{ref}} = 2\,\text{m}^2\,\text{kg}^{-1}\,\text{y}^{-1}$ (bottom row). The other parameter values are fixed as given in Section 2.3. The corresponding N profiles are shown in Figure 4.

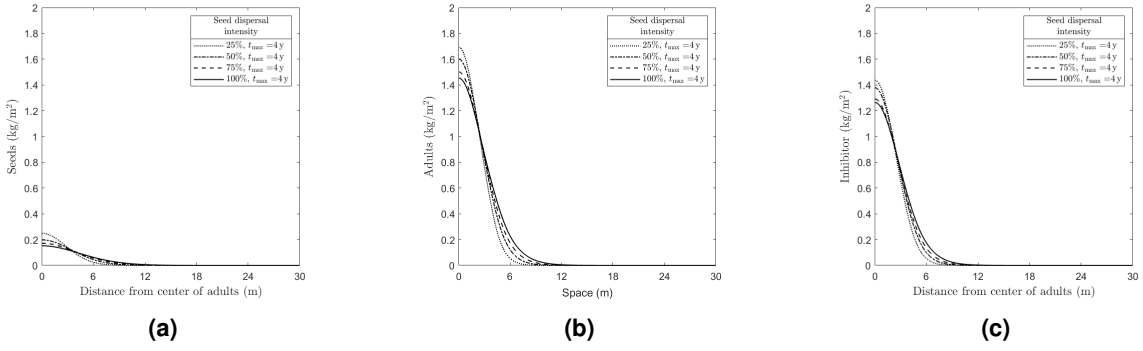

Figure 9. Numerical investigation of the seed dispersal influence on the emergence of Janzen-Connell distributions for S (a), A (b), and I (c). The plots show the variables' profiles at time $t = t_{\max}$ (where the maximum amplitude for N is reached, indicated in the legend) obtained by simulating System (1) for different values of the seed dispersal coefficient d_S, corresponding to different percentages of $d_S^{\text{ref}} = 3\,\text{m}^2\,\text{y}^{-1}$ (the other parameter values are fixed as in Section 2.3). The corresponding N profile is shown in Figure 6.

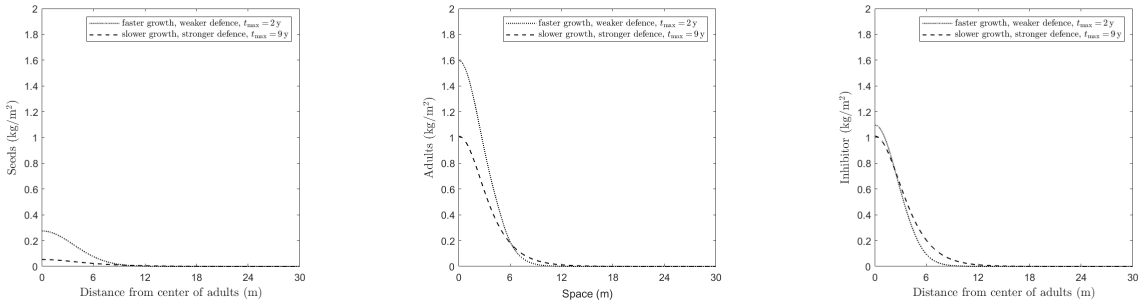

Figure 10. Numerical investigation of the influence of different growth/defence approaches on the emergence of Janzen-Connell distributions for S (a), A (b), and I (c). The plots show the variables' profiles at time $t = t_{\max}$ (where the maximum amplitude for N is reached, indicated in the legend) obtained by simulating System (1) for a species with faster growth, weaker defence ($g_A = 0.9\,\mathrm{y}^{-1}$ and $r_P = 2\mathrm{m}^2\,\mathrm{kg}^{-1}\,\mathrm{y}^{-1}$, dotted line) and slower growth, stronger defence mechanisms ($g_A = 0.02\,\mathrm{y}^{-1}$ and $r_P = 0.1\,\mathrm{m}^2\,\mathrm{kg}^{-1}\,\mathrm{y}^{-1}$, dashed line), respectively (the other parameter values are fixed as in Section 2.3). The corresponding N profile is shown in Figure 6.