Appendix S2. Null model and types of three-dimensional associations

Our work focuses on three-dimensional species-species spatial associations. We estimate the location of individual crowns and who shades who in the Luquillo Forest Dynamic Plot (LFPD) in Puerto Rico. We compare the estimated overlaps between crowns of each pair of species with the overlap expected by a null model. Then we classify the spatial association between each pair of species into four categories: (1) horizontal segregation; (2) horizontal aggregation and vertical segregation; (3) three-dimensional aggregation; or (4) random.

Our null model breaks the spatial (three-dimensional) associations between any two species by randomizing the location of individual trees in the horizontal and the vertical dimensions. Horizontal location of species with respect of each other are broken. Relative species heights are also broken. Diversity, relative abundances, individual crown areas, and spatial aggregation within species, are all kept as observed.

Horizontal randomization

The horizontal randomization involves a torus translation within an area of arbitrary shape. Such translation respects, as much as possible, the horizontal aggregation within species, which may be caused by dispersal limitation or other causes unrelated to species-species interactions or species-level niche preferences. Therefore, our model avoids identifying spurious relationships between species that are caused by conspecific aggregation within both species. This phenomenon is widely studied in the field of ecological/spatial statistics and underlies all works based on torus translations and other techniques.

Fig. S1 contains an example of a fictitious species with individuals uniformly distributed within a square. Note that the translation never takes individuals outside the pre-defined boundaries. With that constraint, the inter-specific aggregation is respected.
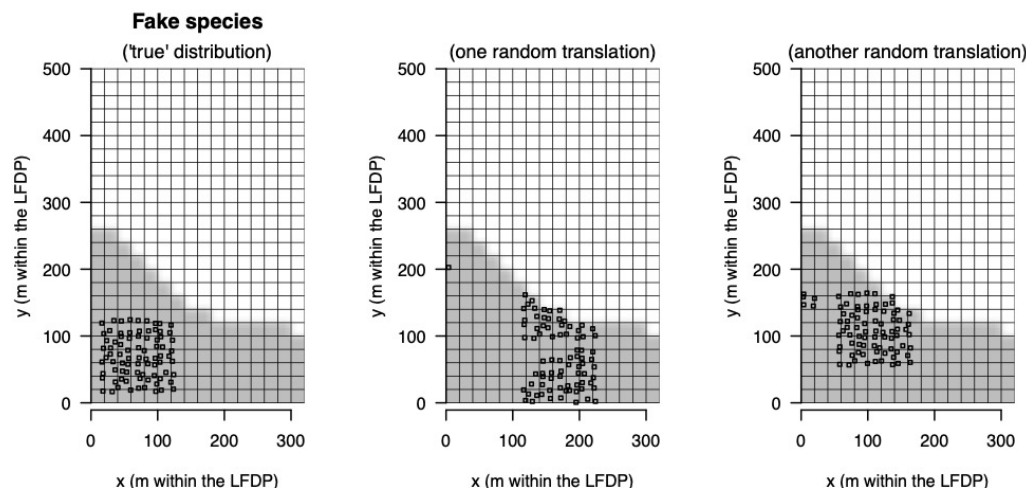

Figure S1. Constrained torus translation of a fictitious species.

In our analysis, we are interested in how different species associate with each other. Therefore, our null model must break spatial associations between species, including horizontal associations. We achieve that by translating each species independently. Fig. S2 contains two examples of two pairs of species whose horizontal association is broken through independent translations. As you can see in the figure, the null model: (1) breaks any horizontal association between each pair of species; (2) respects the horizontal aggregation of individuals within each species (*e.g.* see clusters highlighted by circles and squares); and (3) does not move individuals outside the pre-defined boundaries (be these boundaries habitat types or levels of disturbance or whatever).
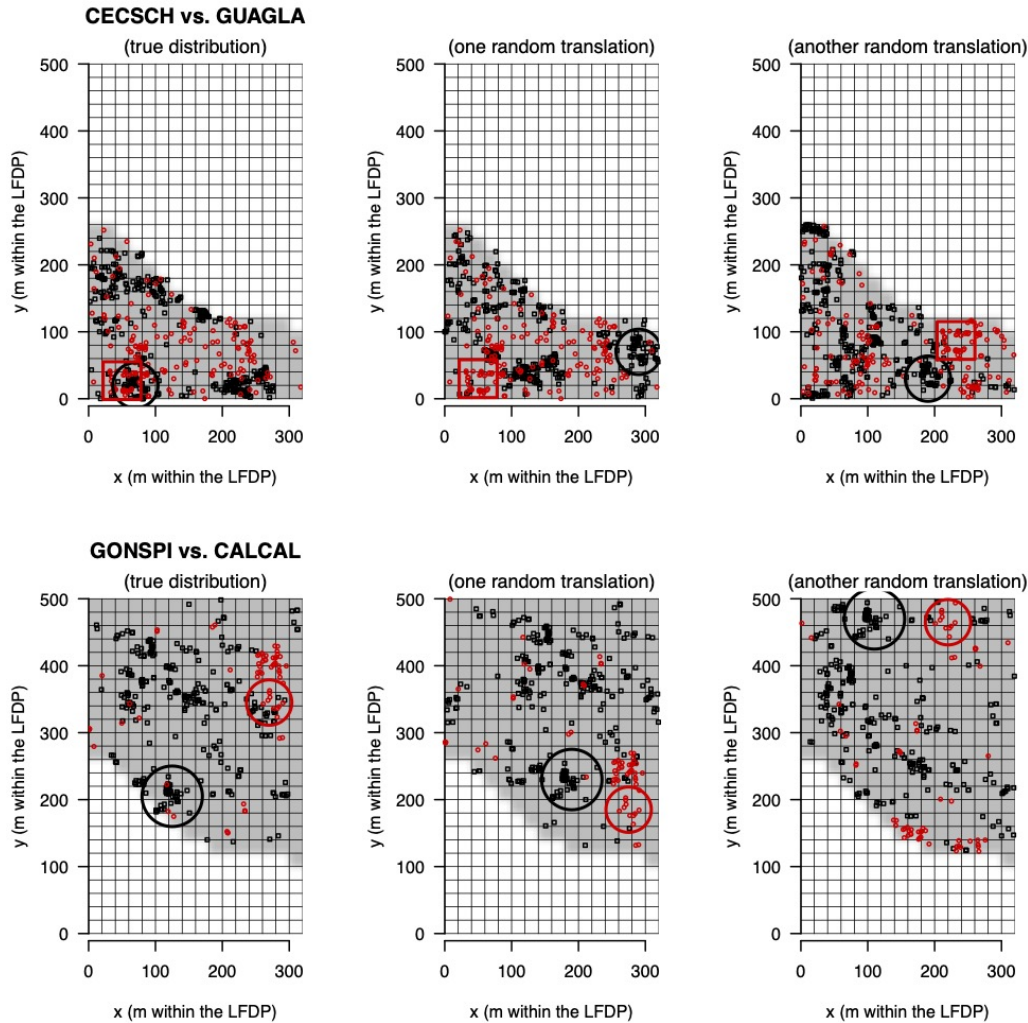

Figure S2. Independent torus translation of two pairs of species within irregular areas. CECSCH = *Cecropia schreberiana*, GUAGLA = *Guarea glabra*, GONSPI = *Gonzalagunia hirsuta*, CALCAL =*Calophyllum brasiliense*.

Vertical randomization

In our null model, the vertical randomization of the location of the individual crowns is a simple permutation of all the heights of all the individuals in the forest (all species combined). This randomization breaks any relationship between the height and crown area of individual trees. Note that differences in species size is one way of organizing species in the vertical space. Larger trees are taller and they occupy higher positions in the system. This vertical stratification is precisely the phenomenon that we want to identify. For that reason, we permuted heights among individuals with all species combined. If we permuted heights exclusively within species, we would be unable to identify vertical organization that results from differences in species sizes.

Crown area remains unchanged

Differences in species size is one way of organizing species in the vertical space, but they should have no role in the horizontal associations between species. For that reason, the crown area remains unchanged in our null model. If crown size was not kept constant when building the null model, small species would have null crowns larger than the observed crowns, and the expectation for them would be to overlap in relatively many squared meters with other crowns. The equivalent reasoning applies for typically large species: if we permuted crown area we would expect (relatively) small areas of overlap. This null model would not be very useful, as we would be interpreting as "horizontal association" what is just inter-specific differences in size.

Categorization of spatial associations between pairs of species

In a forest, species i shades species j in a given amount of squared meters ($A_{i>j}$), and species j shades species i in a different amount, $A_{j>i}$. Without considering who shades who, both overlap $A_{i\leftrightarrow j} = A_{i>j} + A_{j>i}$.

Common species will tend to shade other species in more squared meters in total, while rare species will only have limited areas of crown overlap with others. When species i is consistently taller than j, we will expect $A_{i>j} \gg A_{j>i}$ but, in general, both species can shade each other in similar amounts if they are common and tend to occupy the same horizontal locations. If they avoid each other in the horizontal space they will not shade each other, regardless of their relative heights and abundances ($A_{i\leftrightarrow j} \approx 0$, which implies $A_{i>j} \approx 0$ and $A_{j>i} \approx 0$).

In our work, we compare the observed values of overlap with the expected values of overlap after 999 randomizations using our null model. We describe the deviation between the estimated and the expected overlap between the crowns of each pair of species using a Standardize Effect Size; $SES = (observed - mean(expected))/sd(expected)$. These values are calculated for each pair of species with consideration of who shades who. *I.e.* there is a value for $SES_{i>j}$ and a different value for $SES_{j>i}$. There is also a SES for the overlap between both species without considering who shades who: $SES_{i\leftrightarrow j}$. $SES_{i>j}$ and $SES_{j>i}$ are independent of each other: both can be zero, or low, or one can be low and the other high, or *vice versa*, or both can be high. Their values will depend on how the species organize in the space. $SES_{i\leftrightarrow j}$, in contrast, is not independent from $SES_{i>j}$ and $SES_{j>i}$: when $SES_{i>j}$ and/or $SES_{j>i}$ are low, $SES_{i\leftrightarrow j}$ will be low.

Besides of calculating SES for the observed overlaps between species, we calculate SES for the null values of $A_{i>j}$, $A_{j>i}$ and $A_{i\leftrightarrow j}$ as well. The SES of null values are just the scaled null values: $SES^{null} = (null - mean(null))/sd(null)$. These values will be denoted $SES_{i>j}^{null}$, $SES_{j>i}^{null}$ and $SES_{i\leftrightarrow j}^{null}$. The distribution of $SES_{i>j}^{null}$, $SES_{j>i}^{null}$ and $SES_{i\leftrightarrow j}^{null}$ serve as a reference to define appropriate SES thresholds to keep Type I error rate at a pre-defined level when categorizing the associations between species; this is explained below in detail.

In our work, we classified all associations between each pair of species into four categories:

1. **Horizontal segregation**: two species 'avoid' each other in the horizontal.
2. **Horizontal aggregation and vertical segregation**: two species 'attract' each other in the horizontal but 'avoid' each other in the vertical dimension.

3. **Three-dimensional aggregation**: two species 'attract' each other both horizontally and vertically.
4. **Random association**: two species associate as expected by the null model, both horizontally and vertically.

This categorization is not the only possible; it is simply what we considered useful to group spatial structural patterns in a way comparable to the existing literature. The proposed categories are clearly linked to niche partitioning in one or more niche axes, or none. Below we explore these four cases in detail, included operational definitions, and give some biological interpretation to each of them.

For illustration purposes, we will use simulated 'linear' forests. Fig. S3 shows the lateral projection of the estimated location of 500 crowns in a 300 m x 14.42 m strip in the southwest corner of the LFDP plot in Luquillo.
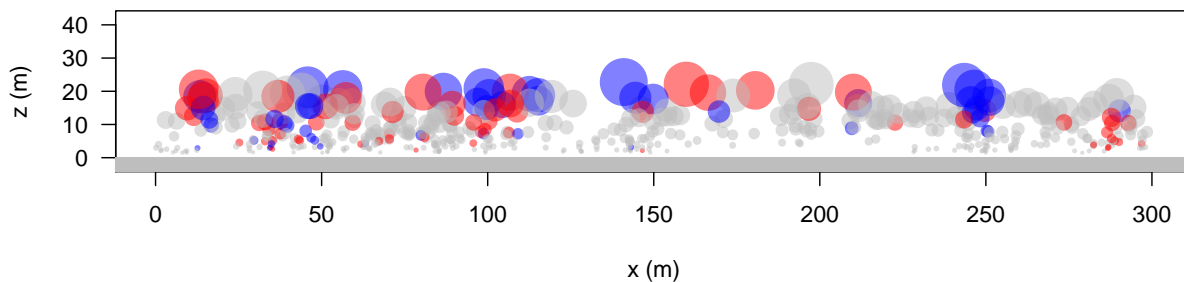

Figure S3. A narrow strip of true forest in Luquillo. Blue = *Dracyodes excelsa*; red = *Manilkara bidentata*. Crown size and location in the z axis (height) are estimated from local allometries. Location in the x axis is taken from the last census in the plot. Trunks are not plotted in these schematic figures to avoid cluttering them.

In this appendix, our simulated forests will be similar to this piece of real forest but perfectly linear, *i.e.* in our simulated scenarios all the N=500 trees will share the same y coordinate, as if located in a perfectly straight line. A torus translation in one dimension x is much simpler and easier to visualize than the difficult torus translation explained above for two horizontal dimensions (x, y). A translation in one dimension assigns new locations equal to $(\mathbf{x} + u)$ mod x_{max}, being $\mathbf{x}$ the vector with the original positions, u a single random number (the "jump" during the translation), and x_{max} the maximum possible value of x ($x_{max} = 300m$ in our case). Although much simpler, it has the same interpretation in terms of breaking spatial relationships between species while respecting the spatial aggregation within any given species. When we apply the horizontal translation plus the vertical permutation the resulting forests looks like those in Fig. S4. We can infer the type of association that any two species have by comparing the overlap between their crowns with the overlap expected in (many) null forests.
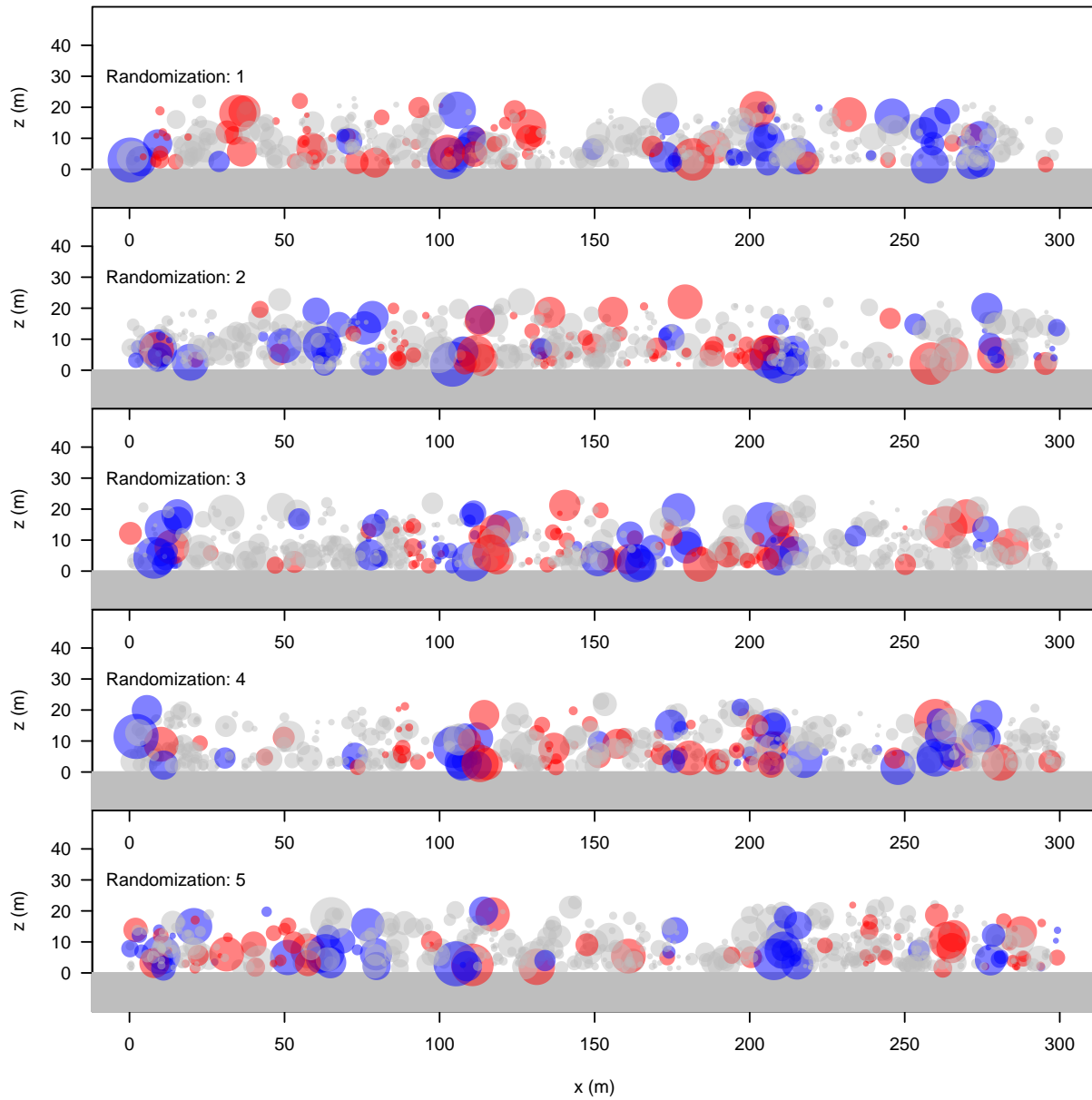
Figure S4. Several randomizations of the same narrow strip of forest within the LFDP plot in Luquillo, Puerto Rico. Blue = *Dracyodes excelsa*; red = *Manilkara bidentata.*

5

Case 1: Horizontal segregation between two species

Two species may have different soil preferences or, for any other reason, appear in different horizontal locations (Fig. S5). This will be reflected in less overlap between their crowns than expected by the null model, regardless of their relative vertical positions (third panel in Fig. S6). Note that the relative position in the vertical dimension is not relevant when two species do not share the same horizontal locations, as they will not shade each other anyway.

Operationally, we looked at the adirectional SES, $SES_{i\leftrightarrow j}$. In particular, we assigned a given spatial association between two species to this class if $SES_{i\leftrightarrow j} < Q(p = 0.05, SES_{i\leftrightarrow j}^{null})$, where $Q(p = 0.05, SES_{i\leftrightarrow j}^{null})$ is the 5% quantile in the distribution of $SES_{i\leftrightarrow j}^{null}$. By this definition, if associations in the empirical forest are similar to the associations in the null forests, we will have a 5% of associations assigned to the "horizontal segregation" category just by chance.
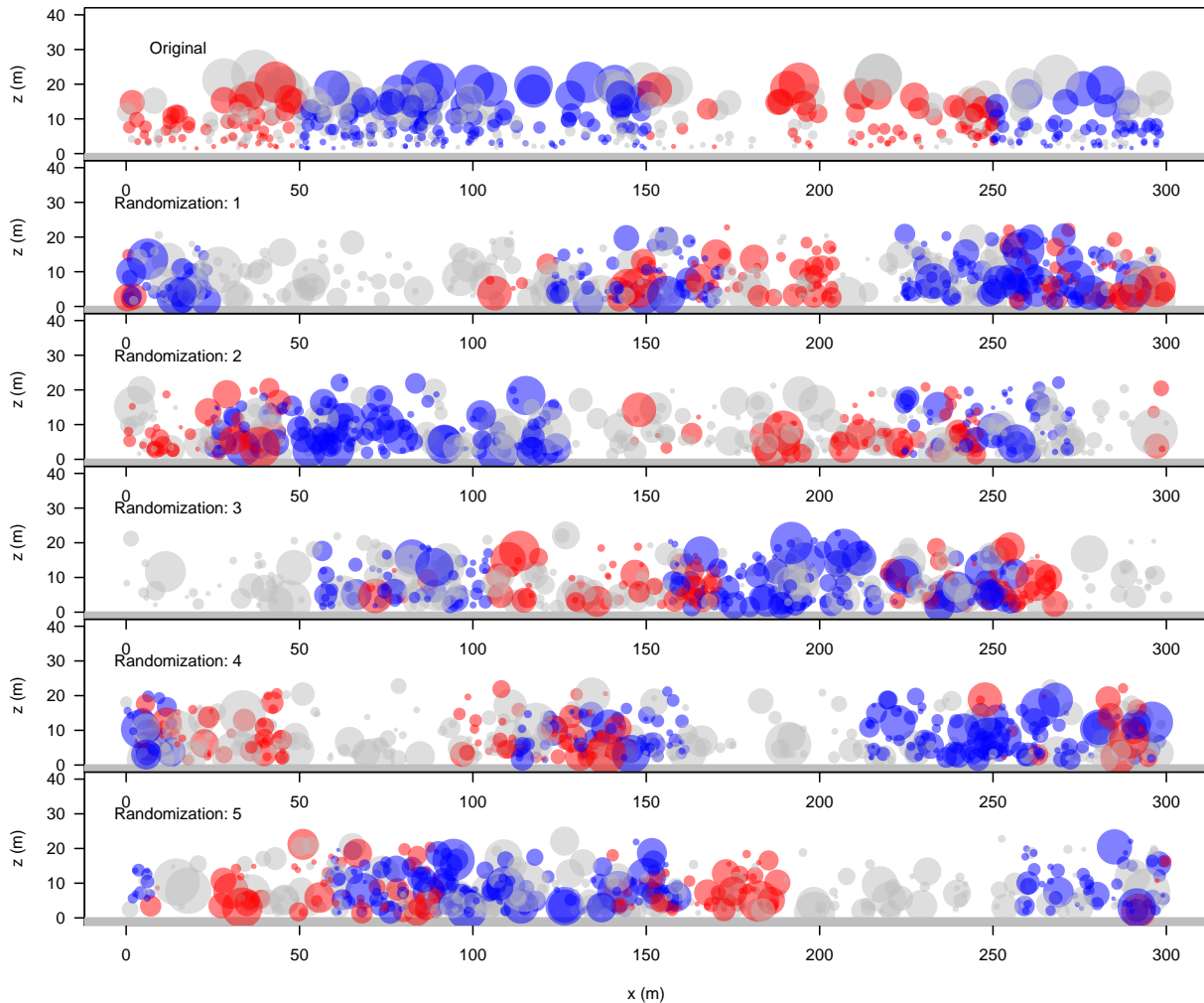

Figure S5. Horizontal segregation between two simulated species (blue and red): original distribution and five examples of a randomized distribution. We can think of both species as soil specialists whose distribution reflects the distribution of different soil types, but without any particular preference towards any forest stratum or vertical location.
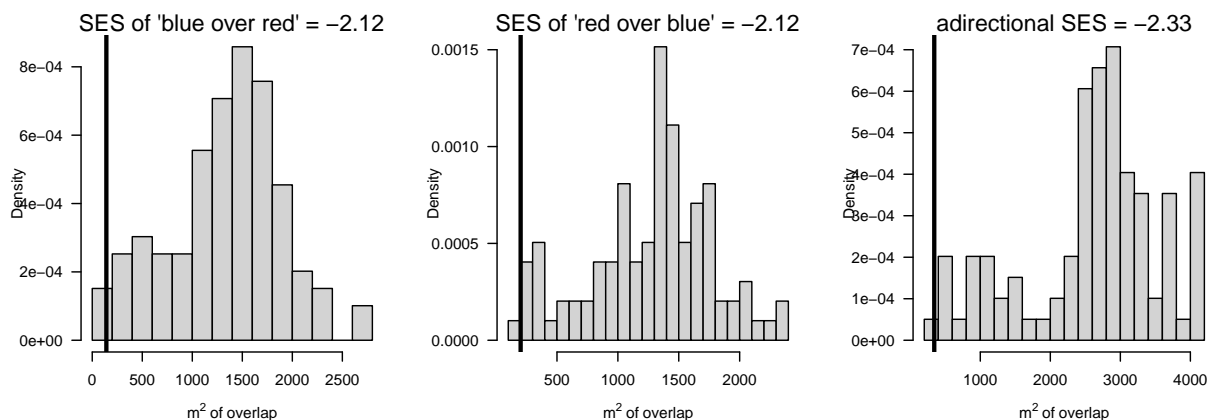
Figure S6. Accumulated overlap between two species that avoid each other horizontally: expected (histogram) vs. observed (vertical line). SES: the corresponding Standardized Effect Size. For the study of horizontal segregation only the third panel (adirectional SES) is relevant.

Case 2: Horizontal aggregation and vertical segregation between two species

Two species can occupy the same horizontal locations (*e.g.* because both prefer a certain soil type) and, at the same time, partition the light environment (Fig. S7). The expectation is that species i will shade the species j more than expected by chance, while the species j will shade species i less than expected by chance; or *vice versa* (Fig. S8). This would be a classical case of multidimensional niche partitioning where preference towards different strata in the forest can facilitate coexistence of two species that share preference towards the same horizontal conditions or locations.

Operationally, we assigned a given association to this category if $SES_{i>j} > Q(p = v_{ij}, SES_{i>j}^{null})$ *and* $SES_{j>i} < Q(p = 1 - v_{ij}, SES_{j>i}^{null})$, where v_{ij} is a value between 0 and 1 chosen in a way that, for this particular pair of species i and j, only 5% of null associations exceeded both thresholds simultaneously. In other words, we assigned an association to this category only when $SES_{i>j}$ was very high and $SES_{j>i}$ very low simultaneously, adjusting symmetrical thresholds to not to exceed a 5% Type I error rate, as in the previous category. By this definition, if associations in the empirical forest are similar to the associations in the null forests, we will have a 5% of associations assigned to the "horizontal aggregation and vertical segregation" category just by chance.
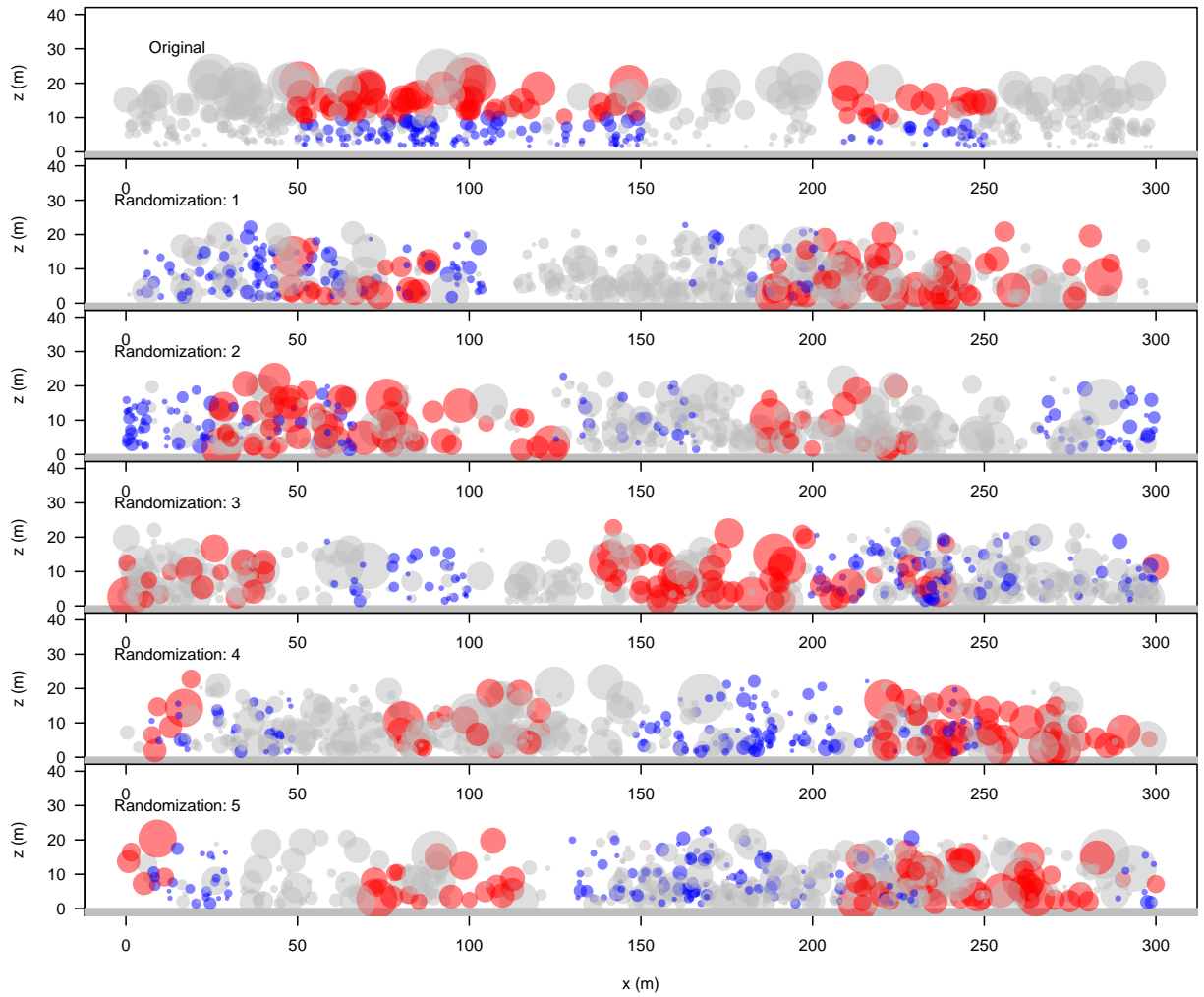
Figure S7. Horizontal aggregation and vertical segregation between two simulated species: a canopy species (red) and an understorey species (blue) that tend to occupy the same horizontal locations. The figure contains the original distribution and five examples of a randomized distribution based on our null model.
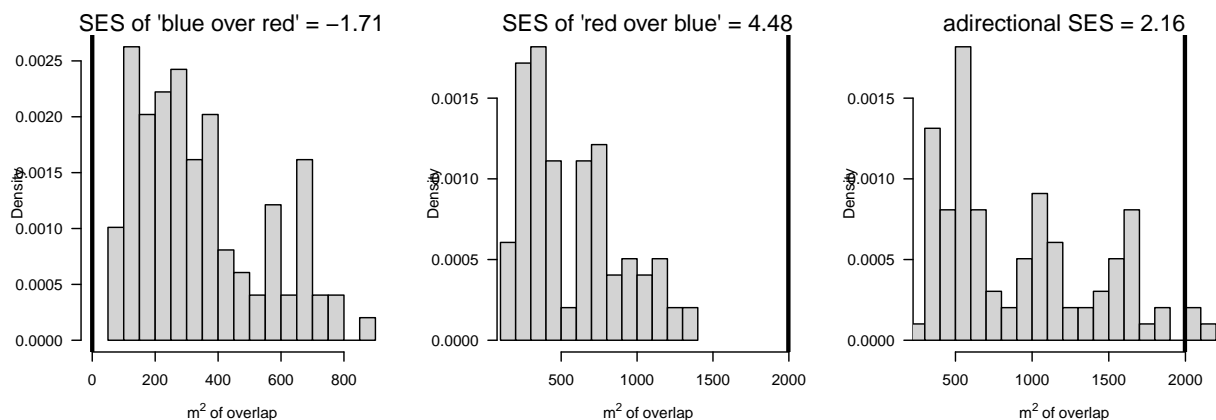
Figure S8. Accumulated overlap between two simulated species that avoid each other vertically: expected (histogram) vs. observed (vertical line). SES: the corresponding Standardized Effect Size. For the study of simultaneous horizontal aggregation and vertical segregation only the two first panels (directional overlaps) are relevant. The adirectional overlap (third panel), of course, also informs about the horizontal aggregation component.

Case 3: Three-dimensional aggregation between two species

This case involves direct competition both for horizontal (soil) niches and vertical (light) niches (Fig. S9). When two species compete for exactly the same locations in the three-dimensional space, some individuals of i will shade individuals of j and some individuals of j will shade individuals of i. For this to be a meaningful pattern, they must shade each other more than the expected by chance (Fig. S10).

Operationally, we assigned a given association to this category if $SES_{i>j} > Q(p = w_{ij}, SES_{i>j}^{null})$ and $SES_{j>i} > Q(p = w_{ij}, SES_{j>i}^{null})$, where w_{ij} is a value between 0 and 1 chosen in a way that, for this particular pair of species i and j, only 5% of null associations exceeded both thresholds simultaneously. In other words, we assigned an association to this category only when $SES_{i>j}$ and $SES_{j>i}$ were high simultaneously, using the same relative threshold to not to exceed a 5% Type I error rate. By this definition, if associations in the empirical forest are similar to the associations in the null forests, we will have a 5% of associations assigned to the "three-dimensional aggregation" category just by chance.
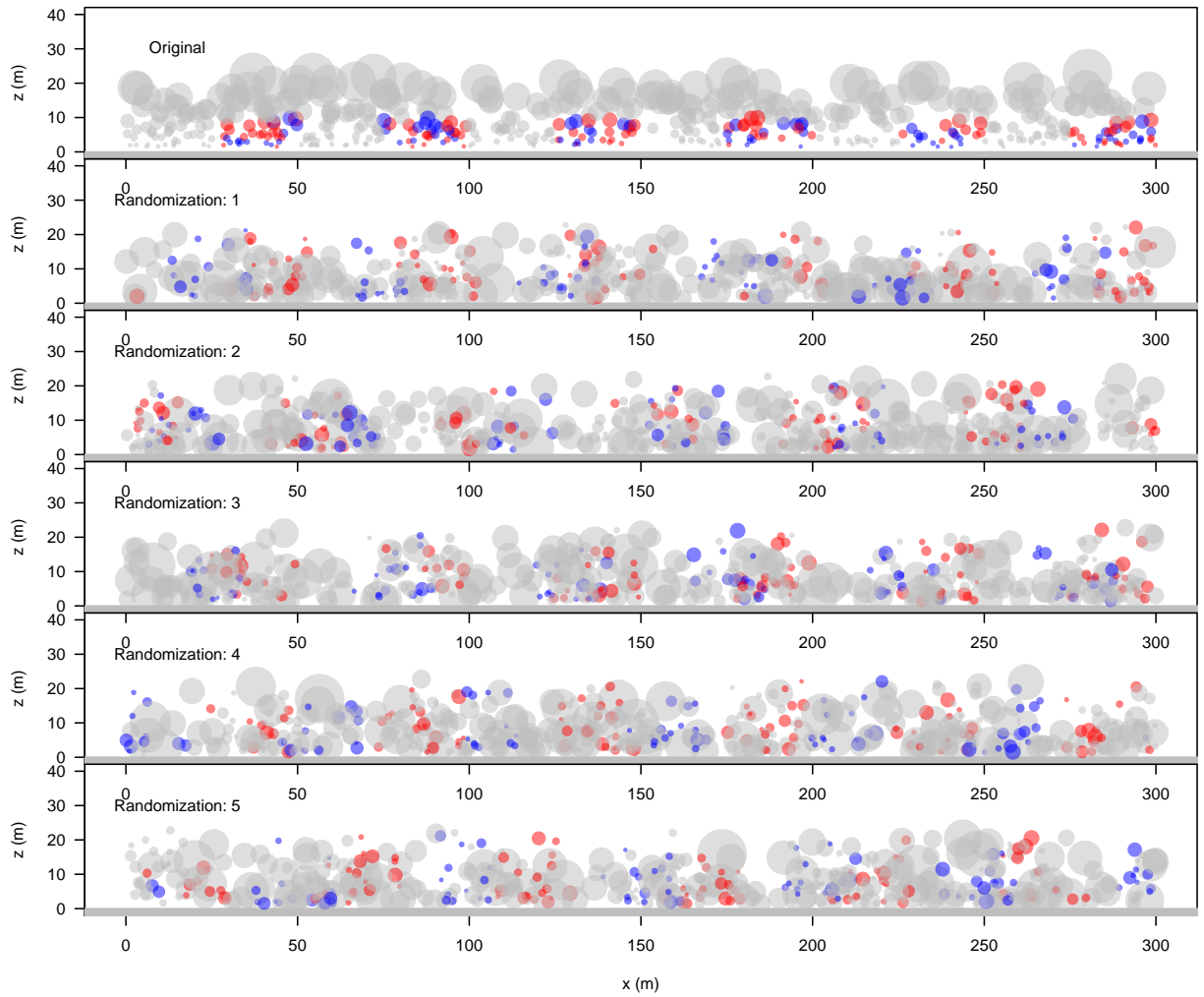
Figure S9. Three-dimensional aggregation between two simulated species: both are understorey species of similar height (red and blue) that tend to occupy the same horizontal locations as well. The figure contains the original distribution and five examples of a randomized distribution according to our null model.
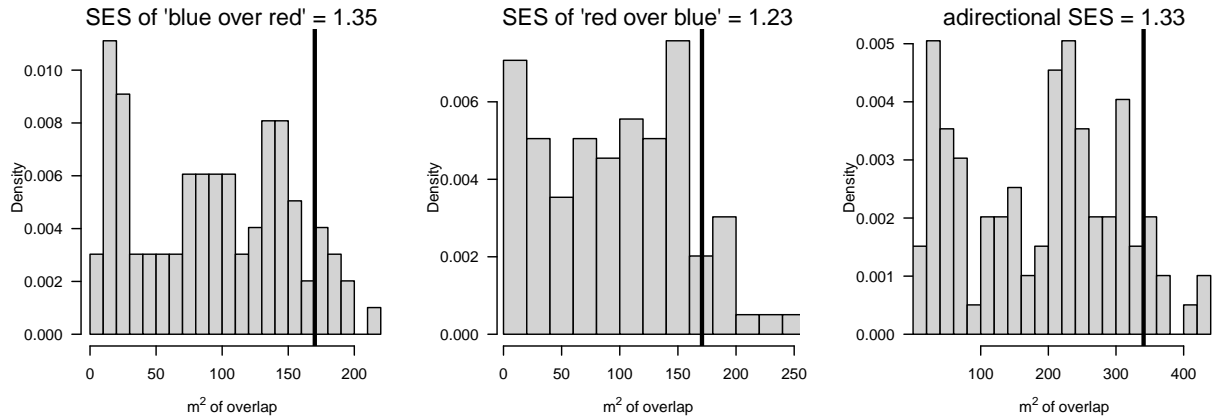
Figure S10. Accumulated overlap between two simulated species that attract each other in three dimensions: expected (histogram) vs. observed (vertical line). SES: the corresponding Standardized Effect Size. For the study of three-dimensional aggregation only the two first panels (directional overlaps) are relevant. The adirectional overlap (third panel) also informs about the horizontal aggregation component.

Case 4: Random association between two species

A random association between two species means that their crowns are located, relative to each other, approximately as in the null communities. In such a forest, trees are completely indifferent to each other, so the presence of one individual does not influence the presence or location of another individual. In these random forests, taller trees have larger crowns but, other than that, the system looks rather disorganized (Fig. S11). The accumulated overlap between the crowns of both species are similar to those expected by chance (Fig. S12). Operationally, we assigned to the "random" category all pairs of species that did not fulfill the conditions for the three categories above. By this definition, if associations in the empirical forest are similar to the associations in the null forests, we will have a 85% of associations assigned to the "random" category just by chance (100% - 5% - 5% - 5%).
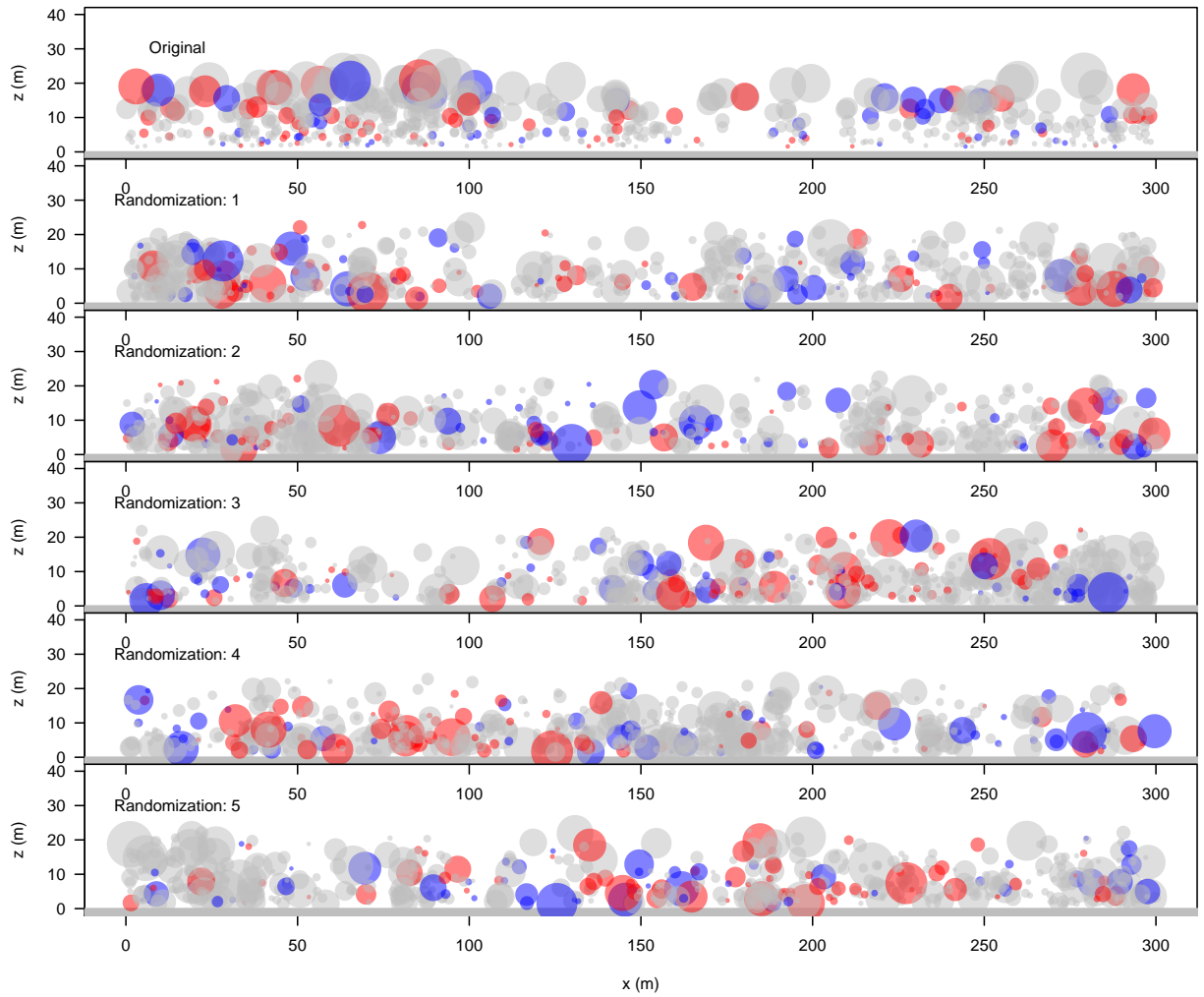
Figure S11. Random association between two simulated species: original distribution and five examples of a randomized distribution.
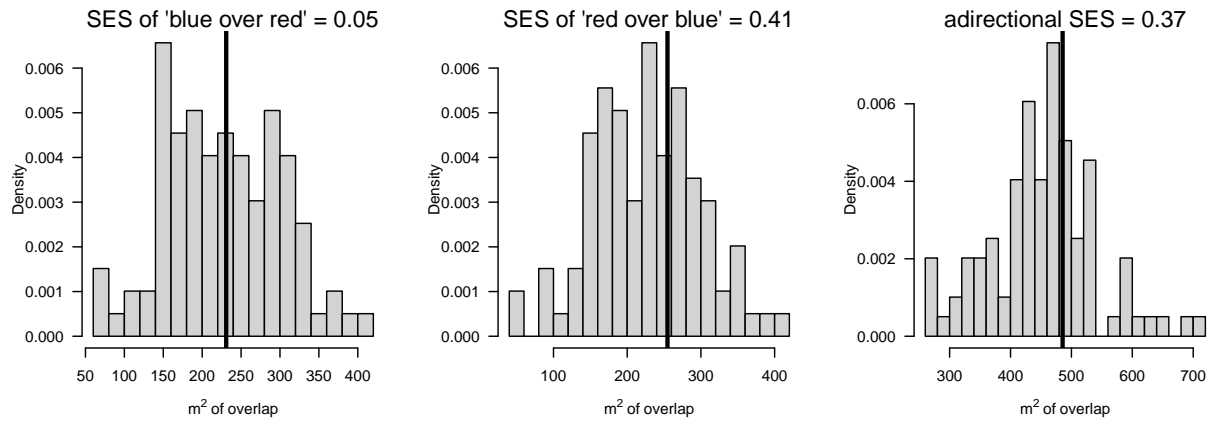
Figure S12. Accumulated overlap between two simulated species that have a random spatial association: expected (histogram) vs. observed (vertical line). SES: the corresponding Standardized Effect Size.