

Stoll, P; Murrell, DJ; Newbery, DM; (2015) Effect sizes and standardization in neighbourhood models of forest stands: potential biases and misinterpretations. Methods in Ecology and Evolution, 6 (10) pp. 1117-1125. 10.1111/2041-210X.12409. Downloaded from UCL Discovery: http://discovery.ucl.ac.uk/1469879

ARTICLE

Effect sizes and standardization in neighbourhood models of forest stands

Peter Stoll^{1,*}, David J. Murrell² and David M. Newbery¹

¹ Institute of Plant Sciences, University of Bern, Altenbergrain 21, CH-3013 Bern,

Switzerland ² Centre for Biodiversity and Environmental Research, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT. UK

Abstract

- 1. Effects of conspecific neighbours on survival and growth of trees have been found to be related to species abundance. Both positive and negative relationships may explain observed abundance patterns. Surprisingly, it is rarely tested whether such relationships could be biased or even spurious due to transforming neighbourhood variables or influences of spatial aggregation, distance decay of neighbour effects and standardization of effect sizes.
- 2. To investigate potential biases, we simulated communities of 20 identical species with log-series abundances but without species-specific interactions. We expected no relationship of conspecific neighbour effects on survival or growth with species abundance. Survival and growth of individual trees was simulated in random and aggregated spatial patterns using no, linear, or squared distance decay.
- 3. Regression coefficients of statistical neighbourhood models were unbiased and unrelated to species abundance. However, variation in the number of conspecific neighbours was positively or negatively related to species abundance depending on transformations of neighbourhood variables, spatial pattern and type of distance decay. Consequently, effect sizes and standardized regression coefficients were also positively or negatively related to species abundance depending on transformation of neighbourhood variables, spatial pattern and distance decay.
- 4. We argue that tests using randomized tree positions and identities provide the best bench marks by which to critically evaluate relationships of effect sizes or standardized regression coefficients with tree species abundance.

Keywords: community dynamics; multiple regression; neighbourhood model; population dvnamics: tropical forest.

1. Introduction

Whether or not conspecific negative density dependence (CNDD) at small neighbourhood scales shapes species abundances in tropical tree communities at larger scales is far from

E-mail: peter.stoll@ips.unibe.ch

resolved and we probably should not even expect the answer to be simple. In principle, there are several possibilities. First, the strength of CNDD is unrelated to abundance. Second, the strength of CNDD is negatively related to abundance (strong CNDD for abundant but weak CDNN for rare species). This would prevent abundant species becoming even more abundant and thereby competitively excluding other species. Moreover, it would confer a rare-species advantage and possibly lead to a community compensatory trend (CCT, Connell et al. 1984). Third, the strength of CNDD is positively related to abundance (strong CNDD for rare but weak for abundant species). This would explain the rarity and low abundance of the species with strong CNDD and the high abundances of species with weak CNDD (Comita *et al.* 2010). There remain though two further possibilities which are that either a mix of positive and negative processes is operating, or the observed relationships are simply spurious (i.e. the result of a statistical artefact).

In an empirical study(Newbery & Stoll (2013) showed negative effects of conspecific neighbours on absolute growth rate (agr) of medium-sized trees. The argument was that reduced growth of an individual tree will – other factors being equal – translate into survivorship reductions and hence affect species abundances. Nevertheless, others argue that effects of conspecifics on survival are more relevant for population dynamics of different species within communities (??). Therefore, the tests reported here simulate individual survival and use a framework of neighbourhood analysis similar to that of Newbery & Stoll (2013) to show that all possible relationships of the strength of CNDD and abundance may emerge without any species-specific or effects of abundances.

Relationships between the strength of CNDD and abundance were investigated using a simple, spatially explicit and individual-based, model which simulated identical species without any species-specific interactions. Thus, any relationships between the strength of CNDD and abundance in communities simulated under these assumptions would not be expected. Nevertheless, relationships do emerge because of interfering effects of spatial patterns and distance decay (i.e. the functional form relating neighbour effects to distance from focal trees, Fig. 1) and, perhaps more importantly, due to the scaling and transforming of the input variables. For example, if rare species have lower variance in the number of conspecifics in their local neighbourhoods compared to common species, scaling is expected to decrease effect sizes (or standardized partial correlation coefficients) of rare relative to common species, possibly leading to spurious negative relationships between the strength of CNDD and abundances. Scaling or standardization is recommended (e.g., Schielzeth 2010) and applied especially in hierarchical Bayesian modelling to speed up or even ensure numerical convergence (e.g., Gelman & Hill 2007).

Motivation to investigate the relationships between the strength of CNDD and abundance more carefully using simulations came from the contrasting outcomes of two recent publications. A consistent *negative* relationship between the strength of CNDD (i.e. effect sizes derived from statistical neighbourhood models) and abundance (total basal area of species) in randomization tests was shown by Newbery & Stoll (2013). But, by contrast, a strong *positive* relationship between the strength of CNDD and abundance was found by Comita *et al.* (2010). Whilst such different results are interesting, and might be explained by different underlying biological mechanisms operating on different species at different locations, before making such a conclusion possible differences arising from the various artefacts of the statistical methods should first be ruled out.

2. Materials and Methods

A completely neutral forest without any species-specific effects was simulated. Initial size distributions of individuals (basal area, *ba*) were log-normal with mean 2 and standard deviation 1. Individuals of 20 identical species with log-series abundances (i.e. 2827, 1408,

935, 699, 557, 462, 395, 344, 305, 273, 248, 226, 208, 192, 179, 167, 157, 147, 139, 132) were placed on plots (200 x 400 m) either randomly or with aggregated spatial patterns. The aggregated pattern was realized by dispersing individuals around 'parent trees' (assigned random locations according to a homogeneous Poisson process), using a Gaussian dispersal kernel with mean 0 and standard deviation 3 m. Thus the species distributions were modeled as a Thomas cluster process, which in turn is a special case of a Neyman-Scott cluster process (Neyman & Scott 1952), and this method means species are spatially independent of one another.

Individual survival was simulated in three steps. First, a linear predictor (y) for survival was simulated for individuals within a border of 20 m using the following multiple regression equation:

$$y = \beta_0 + \beta_1 \log(ba) + \beta_2 \sum_r heterospecificneighbours + \beta_3 \sum_r conspecificneighbours$$
(1)

with β_0 to β_3 the regression coefficients, ba the initial size (basal area) of individuals and the two neighbour terms simply summing the number of heterospecific or conspecific neighbours within a neighbourhood radius (r) of 20 m without taking size or relative size differences between focal individuals and neighbours into account. The regression coefficients were chosen to lead to roughly 50% mortality for each species. Specifically, $\beta_0 =$ 5, $\beta_1 = 2.5$, and $\beta_2 = \beta_3 = -0.05$. Second, the linear predictor was converted to individual survival probabilities using the inverse logit transformation. Third, binomially distributed errors were used to convert the probabilities to the binary variable survival (0's and 1's) by drawing from binomially distributed random numbers. Survival was then used in logistic regressions as the dependent variable. Regressions were run for each species separately. Standardized regression coefficients (b) were obtained from regressions with independent variables standardized by subtracting their mean and dividing by their standard deviation (i.e. scaled). Unlike in the neighbourhood analysis for absolute growth rate (agr) as dependent variable (see below), neighbourhood radii were fixed for the logistic regressions at 20 m. To investigate effects of transformations, the same multiple regression approach as described above, but now with log-transformed neighbour terms was used:

$$y = \beta_0 + \beta_1 \log(ba) + \beta_2 \log(1 + \sum_r heterospecificneighbour) + \beta_3 \log(1 + \sum_r conspecificneighbour)$$
(2)

with β_0 and β_1 as above, and $\beta_2 = \beta_3 = -1.3$. Again, unstandardized (β) and standardized (b) regression coefficients were estimated by logistic regressions (general linear models with binomially distributed error terms). Finally, β_3 and b_3 , as well as variability in numbers of conspecific neighbours within r for each species, were correlated with species abundances (i.e. log[number] of individuals of each species at the plot level).

In the simulations of individual growth, different distance decays and relative size differences were also taken into account because competition is often size- and distance dependent. For each individual, one single growth increment (absolute growth rate, *agr*) was simulated for individuals within a border of 20 m using the following multiple regression equation:

$$\log(agr) = \beta_0 + \beta_1 \log(ba) + \beta_2 \log(1 + \sum_r ba_{HET} / w) + \beta_3 \log(1 + \sum_r ba_{CON} / w) + error$$
(3)

with w = 1 (no distance decay), w = distance (linear distance decay) or w = distance² (squared distance decay, Fig. 1). The neighbourhood terms (ba_{HET} and ba_{CON}) summed the basal areas of bigger heterospecific (HET) or bigger conspecific (CON) neighbours within a neighbourhood radius (r) of 20 m. The random error term was N (0, 0.3). Regression coefficients were $\beta_0 = -0.1$, $\beta_1 = 0.3$ and $\beta_2 = \beta_3 = -0.2$. To verify the simulations, test runs

with random errors set to N (0, 0) were performed. The simulations were realized using C++ (computer code is given in Appendix A of the supplementary material).

Neighbourhood models (as in Stoll & Newbery 2005) were then fitted to the simulated data over all possible combinations of radii for HET and CON neighbours using R (R Development Core Team 2012) and parameter estimates taken from those models yielding the highest adjusted R^2 -values. Five runs with different seeds were performed and estimates of regression coefficients from best fitting neighbourhood models, effect sizes (Cohen 1988; Nakagawa & Cuthill 2007) or standardized regression coefficients (e.g., Warner 2012) averaged across the five runs. Effect sizes (i.e. squared partial correlation coefficients, t²/[t² + residual degrees of freedom], t = t-value) and standardized regression coefficients (b = β 's obtained from regressions with all input variables standardized by subtracting their mean and dividing by their standard deviation) were then correlated with species abundances (i.e. plot level basal area, BA, log transformed). In the case of continuous (as opposed to discrete or binary dependent variables) standardized regression coefficients can also be calculated from unstandardized β 's as b = β * SD_X / SD_Y. A positive correlation of b with abundance implies that less abundant, rare species have stronger CON effects – β is more negative – (as in Comita et al. 2010), whereas a negative relationship implies more abundant species have stronger CON effects (as in Newbery & Stoll 2013). Note, however, that possible correlations of b with abundance may be biased due to correlations of SD_x (or SD_y) with abundance. But if β 's are negative, large SD_x lead to large negative b-values and the relationship with abundance may switch direction not because of a difference in the strength of conspecific interactions between rare and common species, but because of differences in the variability of number or abundance of conspecific neighbours. Moreover, because the simulations and analyses for both survival and agr as dependent variables are based on a multiple regression approach, the basic consequences described above (i.e. possible biases in standardized regression coefficients because of differences in SD_x between rare and common species) are essentially the same independent of the nature of the dependent variable

3. Results

There were no significant regressions for negative conspecific density-dependent effects (regression coefficient β_3 in Eq. 1) on survival and species abundance (Fig. 2) regardless of whether untransformed or log-transformed number of conspecific neighbours were used to quantify neighbourhood. However, variability in number of conspecific neighbours was positively correlated with abundance if untransformed (Eq. 1) but negatively related if log-transformed (Eq. 2). Consequently, standardized regression coefficients were negatively correlated with abundance if number of conspecific neighbours was quantified on the untransformed but positively correlated with abundance if number of conspecific neighbours was quantified on the untransformed but positively correlated with abundance if number of conspecific neighbours were log-transformed. Frequency distributions for rare and common species on untransformed and log-transformed scales (Fig. 3) demonstrate that log-transforming the number of conspecific neighbours for rare species (small values) expands variability but compresses the variability in number of conspecific neighbours for common species (large values). This variability in number of conspecific neighbours increases from rare to common species on transformed scales.

There were no significant regressions for negative conspecific density-dependent effects on growth (regression coefficient β_3 in Eq. 3) and species abundance (plot level basal area) regardless of distance decay or spatial pattern (Fig. 4). Variation in parameter estimates was largest for squared distance decay and random spatial pattern. Best fitting radii for bigger conspecific neighbours were unbiased in neighbourhood models without distance decay and random spatial pattern and with linear distance decay they were slightly underestimated. With estimates (mean \pm SD) of 15.9 \pm 2.6

in the random spatial pattern and 14.5 \pm 3.2, the underestimation was more pronounced with squared distance decay.

Variance in local conspecific neighbour density (within 20 m) varied depending on distance decay and spatial pattern (Fig. 5). A strong negative regression with abundance emerged without distance decay in both spatial patterns. With linear distance decay, the regression was not significant with random spatial pattern but still negative in the aggregated pattern. With squared distance decay, the regression switched to positive in the random pattern, but it was not significant in the aggregated pattern.

As a consequence of variation in local conspecific neighbour density, effect sizes (Fig. 6) and standardized regression coefficients (b_3 , Fig. 7) showed various relations with abundance depending on distance decay and spatial pattern. Without distance decay both effect sizes and standardized regression coefficients were positively related with abundance, regardless of spatial pattern. This was also the case for effect sizes and linear distance decay, whereas standardized regression coefficients were not significantly related with abundance in random spatial pattern but still positively related with abundance in the aggregated pattern. For squared distance decay, both effect sizes and standardized regression coefficients were negatively related with abundance in random spatial patterns but still positively related with abundance in the aggregated pattern. For squared distance decay, both effect sizes and standardized regression coefficients were negatively related with abundance in random spatial patterns but unrelated in aggregated patterns. Apparently, the squared distance decay cancelled the effect of aggregation.

4. Discussion

The simulations and neighbourhood analyses with individual survival or growth as dependent variable showed that estimates of regression coefficients were unrelated to species abundances independent of transformations, spatial pattern and distance decay — as expected based on the simulations of identical species without species-specific interactions. However, variance in local density of conspecifics showed various relationships with species abundances depending on transformations of neighbourhood variables, degree of spatial pattern and form of distance decay. As a consequence, relationships between effect sizes, or standardized regression coefficients, and species abundances were either non-significant, positive or negative.

If untransformed scales are used to quantify conspecific neighbourhoods, relationships with variability and abundance are expected to be generally positive (Fig. B2 in Appendix B) at least in the cases and range of abundances investigated here. In these cases, relationships between standardized effect sizes with abundance will be negative. If, however, log-transformed scales are used to quantify conspecific neighbourhoods, relationships with variance and abundance (Fig. B2) can be modified in all possible ways, i.e. be absent, positive or negative, depending on spatial pattern, exact form of distance decay, but also on whether or not relative size differences are taken into account. There are many and sometimes rather non-transparent possibilities making it very difficult to systematically evaluate the published literature on neighbourhood models and possible relationships between the strength of CNDD and species abundance, especially where details of how variables were handled are incompletely reported, and data have not been archived to allow independent checks.

By using neighbourhood models without distance decay and unstandardized input variables, in single-species analyses, a negative relationship between CNDD and forest-level abundance was found, at least for the first of the two 10-year periods analyzed (Newbery & Stoll 2013). Using no distance decay, yet standardizing before fitting their models, Lin *et al.* (2012) found positive relationships over their dry-season interval. Using an exponential distance decay, Comita *et al.* (2010) centered (subtracted the mean) but did not standardize (divide by standard deviation) their input variables (L. Comita, pers. comm.) and found a strong positive relationship too. Whereas Lin *et al.* (2012) fitted mixed models using

maximum likelihood estimation, i.e. without any prior information being involved, Comita *et al.* (2010) used a hierarchical Bayesian analysis with non-informative priors distributed according to the scaled inverse-Wishart function. This conjugate distribution models the covariance matrix of the species-level regression. Nevertheless, both studies did find positive relationships, thereby apparently supporting one another's conclusions.

The specific scale and distribution of the priors used by Comita *et al.* (2010) might have introduced additional critical information that determined in part the estimation of their coefficients, in a similar way as standardization did in our simulations, and may also have done for Lin *et al.* (2012). Gelman and Hill (2007) discuss the use of the inverse Wishart distribution in some detail, and highlight in particular the need to confirm that Bayesian priors are indeed non-informative across the same ranges of independent variables that result in the posterior probabilities. Dennis (1996) has discussed fundamental issues concerning the use of non-informative priors and Bayesian analysis for ecology in general.

The results of Newbery & Stoll (2013) dealt with effects of conspecific neighbours (as large tree abundance) on growth of small trees, whereas those of Comita *et al.* (2010) concerned conspecific neighbour effects (as either local tree seedling density or tree abundance) on survival of those seedlings. Their result could be more generally important if it were be shown to be fully robust to statistical treatment. It might then support the notion that fundamentally different DD processes are likely operating at the seedling as opposed to the small-tree stage in tropical forest dynamics (Uriarte *et al.* 2004 a, b; Newbery and Stoll 2013).

Because of conceptual similarities of neighbourhood analyses of Comita *et al.* (2010) and those of others (e.g., Uriarte *et al.* 2004a, b; Lin *et al.* 2012), the analysis presented here could be more widely relevant. Since standardization can lead to spurious relationships between CNDD and species abundances, its potential influence needs to be carefully considered when interpreting relationships of small-scale effects of conspecific neighbours on larger scale abundance patterns within diverse tree communities. Similarly, care should be taken when specifying and justifying prior information in hierarchical Bayesian analyses. Our recommendation, following from Newbery and Stoll (2013), is that tests that randomize tree positions and identities indeed provide the best benchmark by which to critically evaluate and judge relationships between effect sizes, or standardized regression coefficients, and tree species abundances.

5. Supplementary Materials

Appendix A

Documented computer code used for the simulations. A detailed description of input parameters and simulation output is provided in the file named growth_files.rtf.

Appendix B

Relationships of variability in local neighbourhoods with species abundances in different competitive scenarios, spatial patterns and various distance decays.

6. References

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences,* Second edn. L. Earlbaum Associates, USA.

- Comita, L.S., Muller-Landau, H.C., Aguilar, S. & Hubbell, S.P. (2010) Asymmetric density dependence shapes species abundances in a tropical tree community. *Science*, **329**, 330-332.
- Connell, J.H., Tracey, J.G. & Webb, L.J. (1984) Compensatory recruitment, growth, and mortality as factors maintaining rain forest tree diversity. *Ecological Monographs*, **54**, 141-164.

Dennis, B. (1996) Discussion: Should ecologists become Bayesians? *Ecological Applications*, **6**, 1095-1103.

- Gelman, A. & Hill, J. (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge.
- Lin, L.X., Comita, L.S., Zheng, Z. & Cao, M. (2012) Seasonal differentiation in densitydependent seedling survival in a tropical rain forest. *Journal of Ecology*, **100**, 905-914.
- Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591-605.
- Newbery, D.M. & Stoll, P. (2013) Relaxation of species-specific neighborhood effects in Bornean rain forest under climatic perturbation. *Ecology*, **94**, 2838-2851.
- Neyman, J. & Scott, E.L. (1952) A theory of the spatial distribution of galaxies. *Astrophysical Journal*, **116**, 144-163.
- R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Schielzeth, H. (2010) Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution,* **1**, 103-113.
- Stoll, P. & Newbery, D.M. (2005) Evidence of species-specific neighborhood effects in the dipterocarpaceae of a Bornean rain forest. *Ecology*, **86**, 3048-3062.
- Uriarte, M., Condit, R., Canham, C.D. & Hubbell, S.P. (2004a) A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? *Journal of Ecology*, **92**, 348-360.
- Uriarte, U., Canham, C.D., Thompson, J. & Zimmerman, J.K. (2004b) A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs*, **74**, 591-614.
- Warner, R.M. (2012) Applied statistics: from bivariate through multivariate techniques. Sage, Los Angeles, USA.

Table 1

Average best fitting radii \pm standard deviation (SD) for bigger conspecifics in neighbourhood models (Eq. 1) across 20 species with identical initial size distributions and log-species abundances and random or aggregated spatial patterns.

	Spatial pattern					
Distance decay	random		aggregated			
no	20.0	±	0.0	19.8	±	0.4
linear	19.6	±	0.5	19.1	±	1.0
squared	15.9	±	2.6	14.5	±	3.2

Figure legends

Figure. 1. Distance decay of neighbourhood effects. In the cut-off model (dashed), the sizes of bigger neighbours with a distance < cut-off are summed. In the linear distance decay (black), the sizes of bigger neighbours are weighed by 1/distance. This is similar to an exponential distance decay (red), which, however, gives somewhat more weight at intermediate distances. A decay of 1/distane² (blue) yields a very rapidly decreasing function. Beyond 20, all three functions give essentially zero weights.

Figure. 2. Regressions of conspecific effects (regression coefficient, β_3 in Eq. 1) on individual survival, variability in number of conspecific neighbours within neighbourhood radius (r = 20), and standardized regression coefficients (b₃) against species abundances (number of individuals at plot level). Twenty species with identical initial size distributions and log-series abundances were simulated in a random spatial pattern. Note that in the panels of the bottom row, the number of neighbours was log-transformed. Data points are means (± 1 SD) from five replicate simulations. The simulated input value of β_3 (dotted lines) was -0.05 (top left) and -1.3 (bottom left). Continuous lines indicate significant (P < 0.05) positive (blue) or negative (red) regressions.

Figure. 3. Frequency distributions of number of conspectic neighbours within neighbourhood radius r (20 m) for individuals of rare and common species in simulated communities with random spatial patterns. Note that the x-axis in the panels of the top row have identical scales. This is also true for the panels of the bottom row. Moreover, the number of conspecific neighbours was log-transformed in the panels of the bottom row. SD indicates the standard deviation of each distribution.

Figure. 4. Regressions of conspecific negative density dependence (regression coefficient, β_3 in Eq. 1) and species abundances (plot level basal area). Twenty species with identical initial size distributions and log-series abundances were simulated without, linear (1/distance) or squared (1/distance²) distance decay of conspecific neighbour effects within 20 m radius in random or aggregated spatial patterns. Data points are means (± 1 SD) from five replicate simulations. The simulated input value of β_3 was -0.2 (green line).

Figure. 5. Regressions of variation in conspecific neighbour density (expressed as SD in basal area of bigger conspecifics, ba_{CON} within 20 m) and species abundances (plot level basal area). Twenty species with identical initial size distributions and log-series abundances were simulated with random or aggregated spatial patterns without, linear (1/distance) or squared (1/distance²) distance decay of conspecific neighbour effects. Data points are means (± 1 SD) from five replicate simulations. Continuous lines indicate significant (P < 0.05) negative (red) or positive (blue) regressions.

Figure. 6. Regressions of effect sizes (squared partial correlation coefficients of β_3 in Eq. 1) and species abundances (plot level basal area). Twenty species with identical initial size distributions and log-series abundances were simulated with random or aggregated spatial patterns without, linear (1/distance) or squared (1/distance²) distance decay of conspecific neighbour effects within 20 m radius. Data points are means (± 1 SD) from five replicate simulations. Continuous lines indicate significant (P < 0.05) positive (blue) or negative (red) regressions.

Figure. 7. Regressions of standardized regression coefficients (b_3) and species abundances (plot level basal area). Twenty species with identical initial size distributions and log-series abundances were simulated with random or aggregated spatial patterns without, linear (1/distance) or squared (1/distance²) distance decay of conspecific neighbour effects within 20 m radius. Data points are means (\pm 1 SD) from five replicate simulations. Continuous lines indicate significant (P < 0.05) positive (blue) or negative (red) regressions.



Distance

FIG. 1.



Abundance (number of individuals, log-scale)









FIG. 4.



FIG. 5.



FIG. 6.



FIG. 7.