

Biodiversity effects and transgressive overyielding

Bernhard Schmid^{1,*}, Andy Hector¹, Prasenjit Saha² and
Michel Loreau³

¹ Institute of Environmental Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

² Physics Institute, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

³ Department of Biology, McGill University, 1205 ave Docteur Penfield, Montreal, Quebec, Canada H3A 1B1

*Correspondence address. Institute of Environmental Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Tel: +41-44-635-5205; Fax: +41-44-635-5711; E-mail: bernhard.schmid@uwinst.uzh.ch

Abstract

Aims

The potential for mixtures of plant species to produce more biomass than every one of their constituent species in monoculture is still controversially discussed in the literature. Here we tested how this so-called transgressive overyielding is affected by variation between and within species in monoculture yields in biodiversity experiments.

Methods

We use basic statistical principles to calculate expected maximum monoculture yield in a species pool used for a biodiversity experiment. Using a real example we show how between- and within-species variance components in monoculture yields can be obtained. Combining the two components we estimate the importance of sampling bias in transgressive overyielding analysis.

Important Findings

The net biodiversity effect (difference between mixture and average monoculture yield) needed to achieve transgressive overyielding

increases with the number of species in a mixture and with the variation between constituent species in monoculture yields. If there is no significant variation between species, transgressive overyielding should not be calculated using the best monoculture, because in this case the difference between this species and the other species could exclusively reflect a sampling bias. The sampling bias decreases with increasing variation between species. Tests for transgressive overyielding require replicated species' monocultures. However, it can be doubted whether such an emphasis on monocultures in biodiversity experiments is justified if an analysis of transgressive overyielding is not the major goal.

Keywords: biodiversity experiments • mixtures • monocultures • overyielding analysis • sampling bias

Received: 28 March 2008 Revised: 28 March 2008 Accepted: 28 March 2008

Introduction

Almost 150 years ago, Darwin and Wallace (1858) summarized the results of the first biodiversity experiments (Hector and Hooper 2002) with the following statement: 'We know that it has been experimentally shown that a plot of land will yield a greater weight if sown with several species and genera of grasses, than if sown with only two or three species.' More than a century later, agronomists compared monocultures with mixtures of two plant species in substitutive replacement series (de Wit 1960). They found that although their so-called relative yield total (RYT) was often larger than one, consistent with some form of 'complementarity', the mixture generally

did not 'transgressively overyield' the better monoculture. Harper (1977) summarized these results with this statement: 'The general conclusion is that there is no advantage to a farmer in sowing a mixture of grass species if his aim is to maximize dry matter production under ideal and constant conditions.' Had he disproved the earlier statement of Darwin and Wallace? As we discuss below, the issue of whether mixtures generally show transgressive overyielding remains a critical issue to this day and lies at the heart of the recent debate over biodiversity experiments.

One key distinction is between overyielding and transgressive overyielding. A mixture overyields when its biomass production is greater than that of the average monoculture of the

species contained in the mixture. This is the approach which is commonly used in biodiversity experiments. In contrast, farmers and more recently restoration managers may be interested in finding a potentially best monoculture species that would be as productive as or even more productive than the mixture (Garnier *et al.* 1997). The contrast between these two approaches has received renewed attention with the interpretation of results from biodiversity experiments. A majority of researchers did find that average productivity increased with species richness, in particular in experimental grassland communities (Balvanera *et al.* 2006, Cardinale *et al.* 2006, 2007). However, when mixtures were compared with monocultures of their component species, the best monoculture was generally as productive as the mixture (Cardinale *et al.* 2006, Hector *et al.* 2002). Probably the only well-known exceptions to this general finding are few-species mixtures of grasses and legumes (Harper 1977, Kirwan *et al.* 2007). Cardinale *et al.* (2006) suggested that the general finding (best monoculture \approx mixture) was evidence for a so-called sampling effect (Aarssen 1997, Huston 1997, Tilman *et al.* 1997): the species which was most productive in monoculture must also have dominated the mixture. In the extreme, such an interpretation would require that the mixture had turned into a monoculture of that species. If there were still other species present in the mixture, then either the ‘sampled’ best species did not make up the full yield or the mixture would have to show transgressive overyielding. Because, in none of the cases where the contribution of the different species could still be recognized at harvest, mixtures had become monocultures, the interpretation of Cardinale *et al.* (2006) is difficult to maintain. In fact, in a reassessment by Cardinale *et al.* (2007), using those studies for which an additive partitioning analysis (Loreau and Hector 2001) could be carried out, it turned out that selection or sampling effects were far less general than originally suggested by Cardinale *et al.* (2006).

So, why was transgressive overyielding so infrequent in biodiversity experiments? Several nonmutually exclusive hypotheses can explain this observation. Here we present one hypothesis, which is very likely to have affected this finding irrespective of other, biological mechanisms. Unequal sampling is inherent to the comparison of a single high-diversity mixture with many or all of the constituent species in monoculture. By comparing the best of many species in monoculture with a single high-diversity mixture of the same species, one tests not only the sampling effect but also the probability of finding an extreme value among a large sample of monocultures that matches the value for the mixture. We will show that this probability depends, first, on the true variation in monoculture yields ‘between’ species and, second, on the sampling variation of monoculture yields ‘within’ species. We start with the first, go on to the second and then combine the two. Once these sources of variation between monoculture yields have been estimated, we will discuss the consequences for analyses of transgressive overyielding in biodiversity experiments.

Variation in Monoculture Yields between Species

Let us assume a farmer can select one out of $N = 10$ plant species whose true monoculture yields (aboveground biomass production during one yearly vegetation period) m_i ($i = 1 \dots N$) follow a normal distribution with a mean of $M = 530 \text{ g m}^{-2}$ and a between-species standard deviation of $S = 238 \text{ g m}^{-2}$ (these values are taken from real data presented below). Which yield m_{\max} could he expect when he plants the highest yielding species? To answer this question, we simply need to find out in which region of the normal distribution the most extreme 10% of all values fall. For a standard normal distribution with mean zero and standard deviation one, 10% of all values are expected to be ≥ 1.28 (Fig. 1a). The value of 1.28 is the x -value for which the area under the standard normal distribution to the left is $p = 1 - (1/N) = 0.9$. This is called the standard normal deviate or quantile of p , $ND(p)$, which

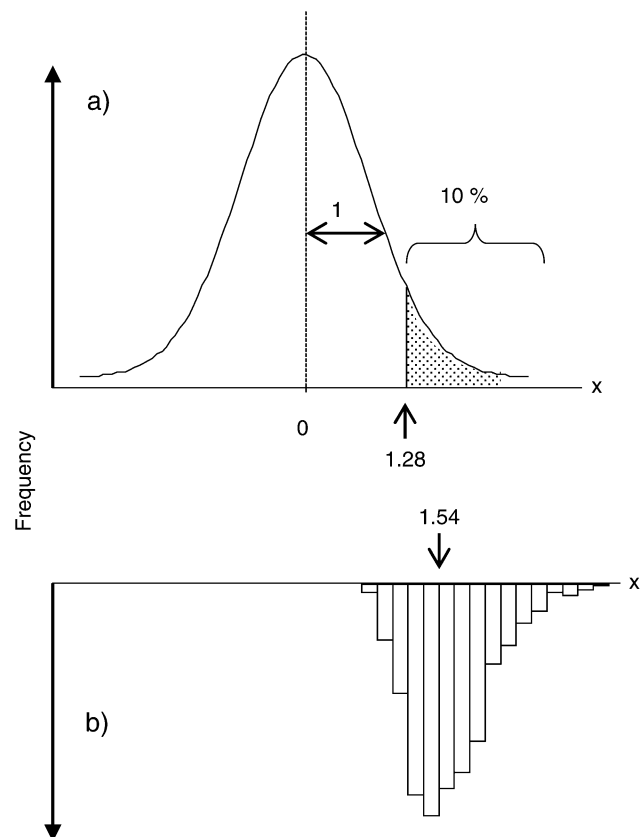


Figure 1 (a) standard normal distribution with mean = 0 and standard deviation = 1. For the normal distribution, 10% of all values lie outside 1.28 times the standard deviation right of the mean. Thus, among 10 values we expect one ≥ 1.28 standard deviations larger than the mean. The value 1.28 is the normal deviate or quantile calculated for the probability $1 - 1/N$ as argument, where N is the number of species in the example from the text. (b) Simulations show that on average the most extreme in a sample of 10 values from a normal distribution lays 1.54 standard deviations above the mean.

can easily be calculated, e.g. in Excel with the function ‘NORMINV’ or in the statistical software R with the function ‘qnorm’. Thus,

$$ND(p) = ND\left(1 - \frac{1}{N}\right) = 1.28.$$

In this example, we would expect one out of 10 values to be at least 1.28 between-species standard deviations larger than the mean:

$$E(m_{\max}) \geq 530 \text{ g m}^{-2} + 1.28 \times 238 \text{ g m}^{-2} = 835 \text{ g m}^{-2}.$$

The average value of m_{\max} obtained by simulation (Fig. 1b), however, is larger than the value obtained from the normal deviate (Script S1 in online supplementary material):

$$\text{Mean}(m_{\max}) \approx 530 \text{ g m}^{-2} + 1.54 \times 238 \text{ g m}^{-2} = 897 \text{ g m}^{-2}.$$

It should be noted that the distribution of simulated values is skewed; thus, the most likely value we would expect to obtain will be between the lower limit of 835 g m^{-2} and the upper value of 897 g m^{-2} . Nevertheless, we can conclude that a mixture of our 10 species would have to produce about 897 g m^{-2} biomass per year in order to reach the same yield as the best species in monoculture. In other words, the mixture would have to produce $d = 897 \text{ g m}^{-2} - 530 \text{ g m}^{-2} = 367 \text{ g m}^{-2}$ more biomass than the average monoculture in order to reach the best monoculture (Fig. 2a). Beyond this ‘break-even point’ the mixture would transgressively overyield the best monoculture.

For the farmer the above result implies that if he has a large number of species with considerable yield differences in

monoculture, he will be better off selecting the best monoculture unless the mixture of all species has a much higher yield than the average monoculture. For a researcher the result implies that transgressiveoveryielding may be found more easily if a pool of species with similar rather than vastly different yields in monoculture is selected for a biodiversity experiment. This is normally taken as a rule for replacement-series experiments (de Wit 1960, Harper 1977, Kirwan *et al.* 2007).

In biodiversity experiments, the between-species standard deviation S can be estimated from the between-species variance component in monoculture yields. This requires, however, that monocultures of species are replicated and that within-species variation in monoculture yields can be pooled across species, a common practical assumption in analysis of variance (ANOVA). Using monoculture data from the second year of the Jena Experiment in Germany (Roscher *et al.* 2004, 2005), we estimated the variance component between 47 well-established species in monoculture as 238 g m^{-2} per year (Table 1), the value used for the above first example. The mean yield of species in monoculture was 530 g m^{-2} , which served as value for M in the first example.

One might argue that the assumption of normally distributed true monoculture yields of species is rarely fulfilled. Like in other biological situations, it is impossible to find negative values and very unlikely to find very large values of monoculture yields. However, even in this case, the normal distribution can still be a useful approximation for calculation purposes. First, species differ in many traits, creating the typical situation of multiple factors influencing a variable, in this case

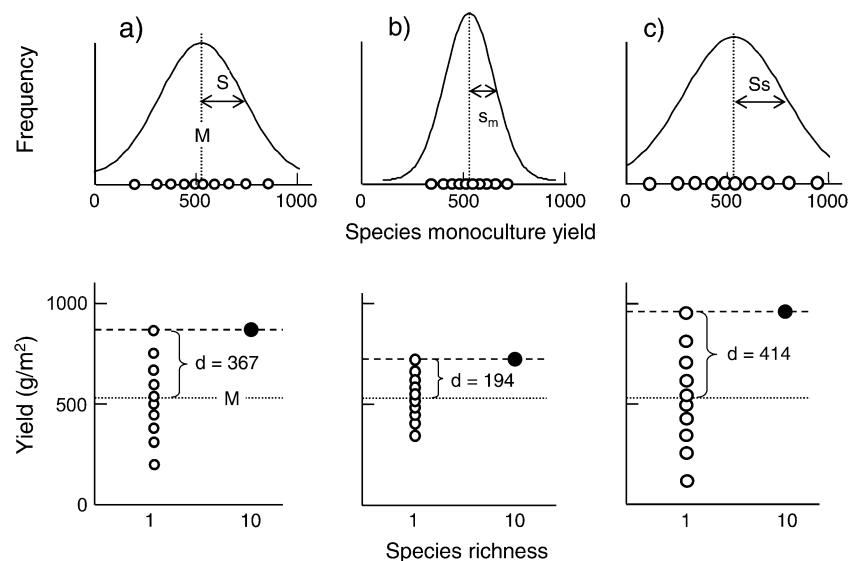


Figure 2 samples of monoculture means of $N = 10$ species with overall mean $M = 530 \text{ g m}^{-2}$ and $n = 2$ replicates per species but different between- (S) and within-species (s) standard deviations: (a) $S = 238 \text{ g m}^{-2}$ and $s = 0 \text{ g m}^{-2}$ (first example in text); (b) $S = 0 \text{ g m}^{-2}$ and $s = 178 \text{ g m}^{-2}$, yielding a standard deviation between species of $s_m = 126 \text{ g m}^{-2}$ (second example in text); (c) $S = 238$ and $s = 178$, yielding a standard deviation between species of $S_s = 269 \text{ g m}^{-2}$ (third example in text). The graphs in the top row show the distributions of monoculture means and the graphs in the bottom row indicate cases for the three examples where a mixture of the 10 species would just begin to transgressively overyield the best monoculture (see text).

Table 1 ANOVA (a) and variance components estimated by restricted maximum likelihood (b) for monoculture data from a biodiversity experiment (Jena Experiment, see Roscher *et al.* 2004, 2005)

(a) ANOVA				
Source of variation	d.f.	s.s.	m.s.	F-ratio
Species	46	6 658 559	144 751	4.56
Replicates	47	1 490 929	31 722	
(b) Estimated variance components				
Source of variation	var.comp.	s.e.(v.c.)	Standard deviation	
Between species	56 515	15 442	$S = 237.7$	
Within species	31 722	6544	$s = 178.1$	

The monoculture yield was estimated as the sum of two aboveground harvests (cut at a height of 3 cm) in May and August 2003. Each species occurred in two replicate plots as monoculture. The 'Species' term in (a) is highly significant ($P < 0.001$); units of original data and standard deviations are gram per square meter. Abbreviations: d.f. = degree of freedom, s.s. = sum of squares, m.s. = mean squares, var. comp. = variance component, s.e.(v.c.) = standard error of variance component, S = between-species standard deviation and s = within-species standard deviation. Note that restricted maximum likelihood estimates of standard deviations do not depend on sample size.

monoculture yield. Thus, according to the central limit theorem we expect a normal distribution. Second, a normal distribution is typically used in ANOVAs when genetic variation between species or lower units is estimated such as in quantitative genetic analyses (Falconer and Mackay 1996). It might be useful in some cases to use log-transformed values of biomass yields because environmental factors often have a multiplicative rather than additive influence on plant growth. In the case of the Jena Experiment, the monoculture means did show a tendency toward a log-normal distribution (Fig. 3); however, we used the normal distribution for simplicity and comparability with biodiversity studies (see, e.g. Cardinale *et al.* 2006) in our demonstration examples. Deviations from a normal or log-normal distribution of species monoculture yields will cause under- or overestimation of the expected maximum monoculture yield.

Variation in Monoculture Yields within Species

As suggested above, one could expect that transgressive overyielding should be more easily detected if all the species in a mixture had the same (true) yield in monoculture. However, yield estimates in monocultures will always have some sampling variation. In this case, the observed species' yields in monoculture will differ even if the true values are all the same. The extent of this sampling bias decreases with increasing number of replicates for each species grown in monoculture and can be estimated if such replication is not totally lacking (as is unfortunately the case in most biodiversity experiments). Thus, let us assume a true monoculture yield of $M = m_i = 530 \text{ g m}^{-2}$ for all species (i.e. $S = 0$) and a within-species sampling

variation of $s = 178 \text{ g m}^{-2}$ (values again taken from the Jena Experiment, see Table 1). Which yield $m_{\text{max obs}}$ would we expect to observe in a biodiversity experiment where each monoculture is replicated once, i.e. $n = 2$ per species (as in the Jena Experiment)?

According to statistical theory, the standard error of the mean for the observed monoculture yields is inversely proportional to the square root of the number of replicates. In this second example given here

$$s_m = \frac{s}{n^{0.5}} = \frac{178}{2^{0.5}} = 126 \text{ g m}^{-2}.$$

In other words, we will obtain the following estimates for M and S :

$$E(M) = \text{mean}(m_i) = 530 \text{ g m}^{-2}, \quad E(S) = s_m = 126 \text{ g m}^{-2}.$$

Applying the normal deviate function again we would expect one out of 10 observed means to be at least $1.28 s_m$'s larger than the mean:

$$E(m_{\text{max obs}}) \geq 530 \text{ g m}^{-2} + 1.28 \times 126 \text{ g m}^{-2} = 691 \text{ g m}^{-2}.$$

If $m_{\text{max obs}}$ is determined by simulation with the values given in the second example, we find an average value of:

$$\text{Mean}(m_{\text{max obs}}) \approx 530 \text{ g m}^{-2} + 1.54 \times 126 \text{ g m}^{-2} = 724 \text{ g m}^{-2}.$$

To test for transgressive overyielding we would normally compare this highest observed species yield in monoculture (a mean over all the replicate monoculture plots of this species) with the mixture yield, mix. Thus, only if $\text{mix} > 724 \text{ g m}^{-2}$ would we start to observe transgressive overyielding. In other words, the mixture would still have to produce about $d = 724 \text{ g m}^{-2} - 530 \text{ g m}^{-2} = 194 \text{ g m}^{-2}$ more biomass per year than the average monoculture to reach the highest observed monoculture mean (Fig. 2b), even though we know from our assumption that the true monoculture means of all species were the same.

This example shows that one should not do a transgressive overyielding analysis before one has tested if species differ in monoculture yields by an ANOVA as shown in Table 1. If the ANOVA does not show differences between species in monoculture yields, it is not appropriate to use the species with the highest observed yield in monoculture for the calculation of transgressive overyielding, rather one should then conservatively use the average or not calculate transgressive overyielding at all. However, to our knowledge monoculture yields have never explicitly been tested for significant differences between species before transgressive overyielding analyses have been carried out in the literature (Cardinale *et al.* 2006, Hector *et al.* 2002). Often this could not be done because there were no replicates for species grown in monoculture. Once a test like the one in Table 1 has shown significant differences between species in monoculture yields, the danger that they did not in fact differ is only that of making a type I error. Even if there is true variation between species in monoculture yields, the

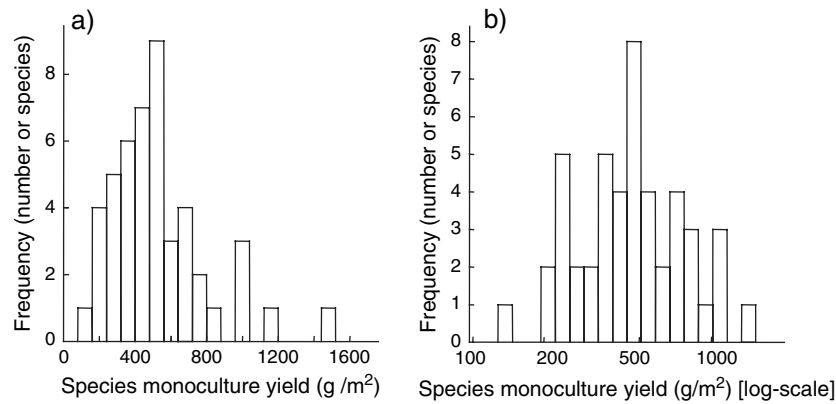


Figure 3 histogram of monoculture means of 47 well-established species from the Jena Experiment (Roscher *et al.* 2004, 2005) on (a) the original and (b) the log scale.

observed monoculture yields will still have a sampling bias. However, as we will show in the next section, in this case the sampling bias has only a weak effect on the analysis of transgressive overyielding.

Variation in Monoculture Yields between and Within Species

In the Jena Experiment a within-species standard deviation of $s = 178 \text{ g m}^{-2}$ combines with a between-species standard deviation of $S = 238 \text{ g m}^{-2}$ (see Table 1). We can guess that the large between-species variation reduces the effect of sampling bias, because it is unlikely that the most extreme within-species deviation from a species mean combines with the most extreme between-species deviation from the overall mean. This is an effect of pooling of variances, which is just the inverse to the partitioning of variances in an ANOVA. The pooled standard deviation S_s is obtained as

$$S_s = \left(S^2 + \frac{s^2}{n} \right)^{0.5}.$$

Thus, combining the first and second example from above, we have $N = 10$, $n = 2$, $M = 530 \text{ g m}^{-2}$, $S = 238 \text{ g m}^{-2}$, $s = 178 \text{ g m}^{-2}$ and

$$S_s = \left(238^2 + \frac{172^2}{2} \right)^{0.5} = 269 \text{ (all in units of } \text{g m}^{-2}\text{)}.$$

Using again the normal deviate for $p = 1 - 1/N = 0.9$ we obtain

$$E(m_{\text{maxobs}}) \geq 530 \text{ g m}^{-2} + 1.28 \times 269 \text{ g m}^{-2} = 874 \text{ g m}^{-2}.$$

If $m_{\text{max obs}}$ is determined by simulation, the value is

$$\text{Mean}(m_{\text{maxobs}}) \approx 530 \text{ g m}^{-2} + 1.54 \times 269 \text{ g m}^{-2} = 944 \text{ g m}^{-2}.$$

In a biodiversity experiment, the mixture of the 10 species would have to produce about $d = 944 \text{ g m}^{-2} - 530 \text{ g m}^{-2} =$

414 g m^{-2} more biomass per year than the observed mean of the best monoculture to reach transgressive overyielding (Fig. 2c). This d -value is only $(414 - 367)/367 \times 100\% = 12.8\%$ larger than the d -value of the first example. Thus, we see that the sampling bias is not as severe if there is true variation between species in monoculture yield as if there is none. This is even the case if, as in the example, s is almost as large as S and within-species replication is minimal with $n = 2$.

This result shows that if a pool of species with vastly different yields in monoculture is selected for a biodiversity experiment, the sampling bias is reduced to a small level where it may be neglected. Remember, however, that transgressive overyielding in such experiments is probably less likely than in experiments using pools of species with similar monoculture yields (see third paragraph in ‘Variation in Monoculture Yields between Species’). In these cases, the sampling bias is not masked as strongly by the between-species variation in monoculture yields and it is therefore more important to reduce the sampling bias via increased replication of monocultures at the within-species level.

Transgressive Overyielding and the Number of Species in Biodiversity Experiments

In all examples, we have used species pools of size $N = 10$. Obviously, if we have larger species pools, it is more likely that the best species in monoculture will have an even more extreme maximum yield, m_{max} . We simply have to take the corresponding normal deviates and simulated multipliers for each particular level of N to obtain minimum expected or average values for $m_{\text{max obs}}$. A range of values is listed in Table 2.

Table 2 shows that the yield of the best species in monoculture increases at a decreasing rate with increasing number N of species in the species pool. In other words, if the best

monoculture species is compared with the mixture of all species in a pool, the mixture has to produce an increasing amount of biomass with increasing N to reach or even transgressively overyield this monoculture species. In fact, the net biodiversity effect NE (Loreau and Hector 2001), which is the difference between the mixture and the average monoculture species, has to be at least as large as the difference between the best and the average monoculture species:

$$NE_{\text{trans}} > ND \left(1 - \frac{1}{N}\right) Ss.$$

In Fig. 4a this relationship between NE_{trans} and the number of species in the mixture (which here corresponds to the number N of species grown in monoculture) is shown for the case with $n = 2$, $M = 530 \text{ g m}^{-2}$, $S = 238 \text{ g m}^{-2}$, $s = 178 \text{ g m}^{-2}$ and $Ss = 269 \text{ g m}^{-2}$, the values used in the third example above.

If biodiversity–productivity relationships saturated at a low species number, e.g. already at $N = 2$ as suggested by Cardinale *et al.* (2006), then NE would be constant for all mixtures with $N \geq 2$. In this case, the log-ratio between the mixture and the best monoculture, the LR_{max} used by these authors to test for transgressive overyielding, would decline as indicated in Fig. 4b, still assuming the values of the third example. With a constant NE of 414 g m^{-2} (corresponding to the d -value in the third example), mixtures with $N < 10$ would show transgressive overyielding, whereas mixtures with $N > 10$ would show transgressive underyielding (Fig. 4b). A similar picture is obtained if Loreau's (1998) D_{max} index is used instead of LR_{max} . This index measures the difference between the mixture and the best monoculture scaled by the best monoculture (Fig. 4c).

Interestingly, Cardinale *et al.* (2006) did not find a decreasing LR_{max} with increasing N . This may be the strongest evidence found so far that the net biodiversity effect NE continues to increase far beyond $N = 2$. However, these authors also found that LR_{max} not only was constant across increasing values of N but also that LR_{max} was not significantly different from zero, indicating that $NE \approx NE_{\text{trans}}$ for the large number of biodiversity experiments they analyzed. This observation led them to suggest a pervasive sampling effect. As mentioned at the beginning, however, this would require that mixtures essentially reduce to monocultures of the best species and in addition that the RYT remains at 1 (Schmid *et al.* 2002). To our knowledge the joint occurrence of these two conditions has never been

Table 2 standard normal deviate (ND) for probability $p = 1 - 1/N$ and corresponding value of multiplier for standard deviation to obtain the expected differences between maximum and average values in a sample of size N (see text for details)

N	P	ND	Multiplier
1	0	undefined	0
2	0.50	0.00	0.57
3	0.67	0.43	0.85
4	0.75	0.67	1.03
5	0.80	0.84	1.16
6	0.83	0.97	1.27
7	0.86	1.07	1.35
8	0.88	1.15	1.42
9	0.89	1.22	1.49
10	0.90	1.28	1.54
20	0.95	1.64	1.87
50	0.98	2.13	2.25
100	0.99	2.33	2.51
500	0.998	2.88	3.04
1000	0.999	3.09	3.24

The values for the multiplier were derived from simulations with 100 000 runs each (see Script S1 in supplementary online material).

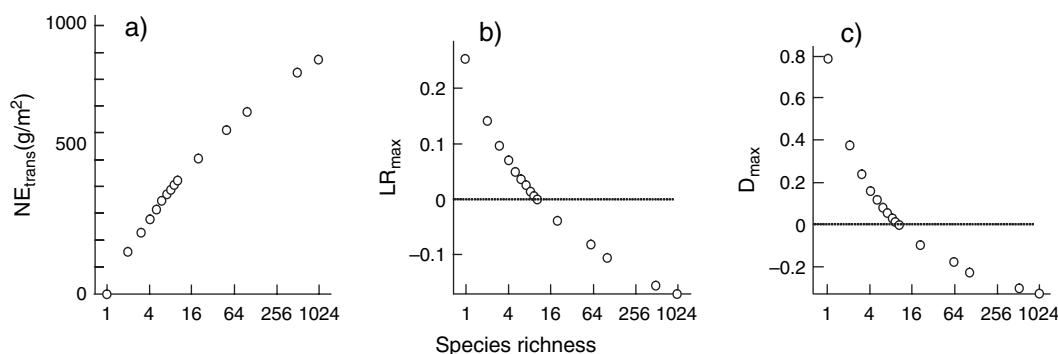


Figure 4 (a) net biodiversity effect NE_{trans} that a mixture of N species must reach to be as productive as the best species in monoculture ($n = 2$, $M = 530 \text{ g m}^{-2}$, $S = 238 \text{ g m}^{-2}$ and $s = 178 \text{ g m}^{-2}$ as in the third example in the text). NE_{trans} is obtained as the difference between the simulated best and the average species in monoculture. If in an experiment a mixture has an observed $NE = NE_{\text{trans}}$, then the measures for transgressive overyield are zero, i.e. $LR_{\text{max}} = D_{\text{max}} = 0$. The points are function values for species richness levels corresponding to the entries in the column headed 'N' in Table 2. (b and c) If mixture yields do not increase with species richness, e.g. if they are constant at $NE = 414 \text{ g m}^{-2}$ (as suggested in the third example in the text), then LR_{max} and D_{max} must decrease with increasing species richness, unless $S = s = 0$.

reported in the literature so far, and new analyses have since rejected this hypothesis (Cardinale *et al.* 2007).

The Danger of Oversampling Monocultures in Biodiversity Experiments

It appears that the main difficulty of current meta-analyses of biodiversity–productivity studies (see, e.g. Balvanera *et al.* 2006, Cardinale *et al.* 2006) lies in the massive oversampling of communities at the lowest level of species richness in most experiments. Ironically, this oversampling is the result of initial concerns about the possibility of so-called sampling effects (Aarssen 1997, Huston 1997, Tilman *et al.* 1997). In fact, this oversampling has now led to sampling effects of the opposite sort, i.e. the chances to find an extreme monoculture value are continuously increasing with the number of species used in an experiment, while at the same time only one mixture of all species at the highest richness level is available. As long as mean values are compared, the problem of unequal sampling at the different richness levels is not severe. But as soon as we take extreme values for comparisons, as is done in analyses of transgressive overyielding and as is sometimes advocated for all analyses of biodiversity experiments (Garnier *et al.* 1997), unequal sampling has severe consequences.

A farmer or restoration manager may want to increase his chances to find an extreme monoculture by using a large ‘diversity’ of species to choose from rather than putting his bets on the ‘single’ mixture of all these species. However, even this approach will only work if the sampling bias due to variation between monocultures within species can be separated from variation between species. This can be done under the assumption of normally (or log-normally) distributed monoculture yields by an ANOVA, if monocultures have been replicated within species—potentially across another experimental factor that was varied such as fertilizer level or soil volume (Dimitrakopoulos and Schmid 2004). Unfortunately, the dogma ‘always grow all species in monoculture’ sometimes can only be fulfilled by reducing the number of replicate monocultures per species. If there is only one monoculture per species, there is of course no possibility to separate the variation between species from the variation within species. And, if we are strict, then there is no possibility to test for transgressive overyielding because we cannot estimate the sampling bias.

There is one additional caveat against choosing a best monoculture over a mixture. If yields of particular species in monocultures or particular species compositions in mixture vary between years, such variation will further increase the sampling bias against the mixture. It is quite likely that the best species in monoculture in one year was just by chance having a good year and will not have as high a yield, or perhaps not even be the best species, in the next year. This increase in the sampling bias will be even stronger if temporal variation is higher in monocultures than in mixtures. Indeed, monoculture yields usually have larger temporal coefficients of varia-

tion than mixture yields (Loreau *et al.* 2001, Pfisterer *et al.* 2004, van Ruijven and Berendse 2007, Tilman *et al.* 2006), which means that they also have larger standard deviations if they reach the same yield (as suggested for the best monocultures). This problem of increased sampling bias due to stronger variation in monocultures is a consequence of the so-called insurance effect (Yachi and Loreau 1999), which has often been mentioned as an important consideration in biodiversity management (Hooper *et al.* 2005) to protect a farmer against low yields in particular years.

Improving the Assessment of Transgressive Overyielding in Biodiversity Experiments

In the face of the above-mentioned problems, are there any possibilities to improve the analysis of previous experiments or to design better experiments with a view to assessing transgressive overyielding? First, from the three examples presented above, it becomes clear that an estimation of the sampling bias in analyses of transgressive overyielding is only possible if replicate monocultures exist for at least some species. If we accept the assumption that the within-species variation in monoculture yield is constant across species, the pooled estimate can be used for all species, including those for which $n = 1$. Second, if there are replicate monocultures within species, we can use an ANOVA to separate the variation in monoculture yields into its between- and within-species components. If this analysis does not reveal significant variation between species, then the average monoculture yield should be taken as the estimate for all species and transgressive overyielding be calculated using this average. However, if there is significant variation between species in monoculture yields, there will be a best species which can be compared with the mixture. The sampling bias then still affects the measure of transgressive overyielding, but it is relatively small, especially if the between-species variance component in monoculture yields is larger than the within-species variance component. Under these conditions, even the minimum number of two replicates for monocultures within species is sufficient.

Supplementary Data

An R-script (Script S1) to run the simulations mentioned in the text with the statistical software R (www.r-project.org) as supplementary data is available at *Journal of Plant Ecology* online.

Funding

German Science Foundation (FOR 456–WE 2618/6-1 to B.S.); Swiss National Science Foundation (31–65224.01 to B.S.); Natural Sciences and Engineering Research Council of Canada (M.L.)

Acknowledgements

We thank three anonymous reviewers and Lin Jiang for useful comments on earlier versions of this paper. We thank Peter Hawthorne, Joe Fargione and David Tilman for sharing their idea about the sampling bias with us and the colleagues from the Jena Experiment for allowing us to use their monoculture data and for comments on the manuscript.

References

- Aarssen LW (1997) High productivity in grassland ecosystems: effected by species diversity or productive species? *Oikos* **80**:183–4.
- Balvanera P, Pfisterer AB, Buchmann N, et al. (2006) Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol Lett* **9**:1146–56.
- Cardinale BJ, Srivastava DS, Duffy JE, et al. (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**:989–92.
- Cardinale BJ, Wright JP, Cadotte MW, et al. (2007) Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proc Natl Acad Sci USA* **104**:18123–8.
- Darwin C, Wallace A (1858) On the tendencies of species to form varieties, and on the perpetuation of varieties and species by natural means of selection. *J Proc Linn Soc Zool* **5**:45–62.
- de Wit CT (1960) On competition. *Versl Landbouwk Onderzoek* **66**:1–82.
- Dimitrakopoulos PG, Schmid B (2004) Biodiversity effects increase linearly with biotope space. *Ecol Lett* **7**:574–83.
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Harlow, UK: Longman.
- Garnier E, Navas ML, Austin MP, et al. (1997) A problem for biodiversity–productivity studies: how to compare the productivity of multi-specific plant mixtures to that of monocultures? *Acta Oecol* **18**:657–70.
- Harper JL (1977) *Plant Population Biology*. London, UK: Academic Press.
- Hector A, Bazeley-White ET, Loreau M, et al. (2002) Overyielding in grassland communities: testing the sampling effect hypothesis with replicated biodiversity experiments. *Ecol Lett* **5**:502–11.
- Hector A, Hooper RE (2002) Darwin and the first ecological experiment. *Science* **295**:639.
- Hooper DU, Chapin FS, Ewel JJ, et al. (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol Monogr* **75**:3–35.
- Huston MA (1997) Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**:449–60.
- Kirwan L, Lüscher A, Sebastia MT, et al. (2007) Evenness drives consistent diversity effects in intensive grassland systems across 28 European sites. *J Ecol* **95**:530–9.
- Loreau M (1998) Separating sampling and other effects in biodiversity experiments. *Oikos* **82**:600–2.
- Loreau M, Hector A (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**:72–6.
- Loreau M, Naeem S, Inchausti P, et al. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**:804–8.
- Pfisterer AB, Joshi J, Schmid B, et al. (2004) Rapid decay of diversity–productivity relationships after invasion in experimental plant communities. *Basic Appl Ecol* **5**:5–14.
- Roscher C, Schumacher J, Baade J, et al. (2004) The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic Appl Ecol* **5**:107–21.
- Roscher C, Temperton VM, Scherer-Lorenzen M, et al. (2005) Overyielding in experimental grassland communities—irrespective of species pool or spatial scale. *Ecol Lett* **8**:419–29.
- van Ruijven J, Berendse F (2007) Contrasting effects of diversity on the temporal stability of plant populations. *Oikos* **116**:1323–30.
- Schmid B, Joshi J, Schläpfer F (2002) Empirical evidence for biodiversity–ecosystem functioning relationships. In: Kinzig A, Tilman D, Pacala S (eds). *Functional Consequences of Biodiversity: Experimental Progress and Theoretical Extensions*. Princeton, NJ: Princeton University Press, 120–50.
- Tilman D, Lehman CL, Thomson KT (1997) Plant diversity and ecosystem productivity: theoretical considerations. *Proc Natl Acad Sci USA* **94**:1857–61.
- Tilman D, Reich PB, Knops JMH (2006) Biodiversity and ecosystem stability in a decade long grassland experiment. *Nature* **441**:629–32.
- Yachi S, Loreau M (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc Natl Acad Sci USA* **96**:1463–8.