Global Change Biology (2009), doi: 10.1111/j.1365-2486.2009.01866.x

Assessing the effect of elevated carbon dioxide on soil carbon: a comparison of four meta-analyses

BRUCE A. HUNGATE*, KEES-JAN VAN GROENIGEN†, JOHAN SIX‡, JULIE D. JASTROW§, YIQI LUO¶, MARIE-ANNE DE GRAAFF‖, CHRIS VAN KESSEL‡ and CRAIG W. OSENBERG**

*Department of Biological Sciences and Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ 86011, USA, †Botany Department, Trinity College, University of Dublin, Dublin 2, Ireland, ‡Department of Plant Sciences, University of California Davis, CA 95616, USA, §Argonne National Laboratory, Biosciences Division, Argonne, IL 60439, USA, ¶Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019-0245, USA, ‖Oak Ridge National Laboratory, Biosciences Division, Oak Ridge, TN 37831-6038, USA, **Department of Zoology, University of Florida, Gainesville, FL 32611-8525, USA

Abstract

Soil is the largest reservoir of organic carbon (C) in the terrestrial biosphere and soil C has a relatively long mean residence time. Rising atmospheric carbon dioxide (CO_2) concentrations generally increase plant growth and C input to soil, suggesting that soil might help mitigate atmospheric CO_2 rise and global warming. But to what extent mitigation will occur is unclear. The large size of the soil C pool not only makes it a potential buffer against rising atmospheric CO_2, but also makes it difficult to measure changes amid the existing background. Meta-analysis is one tool that can overcome the limited power of single studies. Four recent meta-analyses addressed this issue but reached somewhat different conclusions about the effect of elevated CO_2 on soil C accumulation, especially regarding the role of nitrogen (N) inputs. Here, we assess the extent of differences between these conclusions and propose a new analysis of the data. The four meta-analyses included different studies, derived different effect size estimates from common studies, used different weighting functions and metrics of effect size, and used different approaches to address nonindependence of effect sizes. Although all factors influenced the mean effect size estimates and subsequent inferences, the approach to independence had the largest influence. We recommend that meta-analysts critically assess and report choices about effect size metrics and weighting functions, and criteria for study selection and independence. Such decisions need to be justified carefully because they affect the basis for inference. Our new analysis, with a combined data set, confirms that the effect of elevated CO_2 on net soil C accumulation increases with the addition of N fertilizers. Although the effect at low N inputs was not significant, statistical power to detect biogeochemically important effect sizes at low N is limited, even with meta-analysis, suggesting the continued need for long-term experiments.

Keywords: C sequestration, effect size, elevated CO_2, meta-analysis, soil C, statistical power

Received 8 July 2008 and accepted 18 November 2008

Introduction

Soils contain nearly three times the amount of carbon (C) as the atmosphere (Jobbágy & Jackson, 2000; Houghton, 2007), and, on average, C in soils turns over much more slowly than atmospheric carbon dioxide (CO_2). Soil C, and the processes that influence it, affect the CO_2 content of the atmosphere, C sequestration, and climate warming. Environmental changes that influence soil C dynamics could slow atmospheric CO_2 rise and associated warming by promoting soil C storage (e.g. Cramer *et al*., 2001; Johnson & Curtis, 2001), or they could exacerbate warming by causing soil C to decline (e.g. Mack *et al*., 2004; Knorr *et al*., 2005).

Yet, the very properties that make soil C a key reservoir in the global C cycle – namely its large size and slow turnover – also make it difficult to study

Correspondence: Bruce A. Hungate, tel. +1 928 523 0925, fax +1 928 523 7500, e-mail: Bruce.Hungate@nau.edu

Table 1 Summary of four recent meta-analyses on the effect of carbon dioxide (CO_2) on soil carbon (C) accumulation

				Metric interpreted			Cited as support of	
Study	N	Metric analyzed	Weighting function	Relative	Absolute	Nitrogen dependence?	Increased soil C	No effect without N
de Graaff *et al*. (2006)	67	Ln $((E-A)/At+1)$	Time, n	$+1.20\%\ yr^{-1}$	nr^{+}	Greater at high N, factorial test	3	4
Jastrow *et al*. (2005)	35	$\ln(E/A)$	1/Var	+5.6%	$+19\ g\ C\ m^{-2}\ yr^{-1}$	Not tested	15	1
Luo *et al*. (2006)	40	$\ln(E/A)$	1/Var	+5.6%	$+200\ g\ C\ m^{-2}$	Greater at high N, factorial test	12	1
van Groenigen *et al*. (2006)	80	Ln $((E-A)/At+1)$	Time, n; and 1/Var	$+1.17\%\ yr^{-1*}$	nr	Greater at high N across all studies	0	18

N indicates the number of observations included in the meta-analysis. 'Metric analyzed' refers to the effect size metric used in the meta-analysis to determine the significance of the effect, with observations weighted by the 'Weighting function.' By contrast, 'Metric interpreted' shows the summary form of the result used to explain its significance in the text, even though in each case this was not the metric actually analyzed statistically. We argue that metrics of effect size should be identical to expressions that capture the biogeochemical meaning of the result (Osenberg *et al*., 1997, 1999). 'Nitrogen dependence' summarizes the findings about the importance of nitrogen on the response of soil C to elevated CO_2, indicating whether the dependence was tested across all studies or restricted to studies that employed a $CO_2 \times N$ factorial design. 'Cited as support of' describes how citations to date (December 2008 in SciSearch) draw on these studies as support of two contrasting conclusions: elevated CO_2 increases soil C, or elevated CO_2 has no effect on soil C in the absence of added N.

*Weighted average effect size for the three N levels: 0.12% at low N, 2.14% at medium N, and 2.91% at high N.

empirically. Small changes in the size or turnover of soil C are biogeochemically significant but difficult to detect in experiments (Hungate *et al*., 1996; Smith, 2004). Does elevated atmospheric CO_2 increase soil C? By how much and under what conditions? These questions have been the subject of recent meta-analyses of CO_2 enrichment experiments (Jastrow *et al*., 2005; de Graaff *et al*., 2006; Luo *et al*., 2006; van Groenigen *et al*., 2006). Meta-analysis is a useful tool in this case, both because it has the potential to overcome some of the limitations of low statistical power in individual experiments, and because it has the advantage of testing whether responses are general across experiments (e.g. Osenberg *et al*., 1999). These four meta-analyses were conducted at about the same time and, therefore, had access to the same data sources. Thus, they present an opportunity to compare the approaches used in each case and to assess the importance of choices made when applying meta-analysis, choices that are general to applications of the technique in any field (e.g. Englund *et al*., 1999).

These four meta-analyses all concluded that elevated CO_2 increased accumulation of soil C when averaged across all studies. The three meta-analyses that evaluated the role of N fertilization found that the effect of CO_2 on soil C accumulation was greatest when N fertilizer was also added (Table 1). Thus, these four meta-analyses reached similar conclusions on some key points. However, there was less agreement about the magnitude (and significance) of the effect of elevated CO_2 in unmanaged ecosystems (e.g. in the absence of N addition), a point illustrated by citation patterns in the literature: Luo *et al*. (2006) and Jastrow *et al*. (2005) are frequently cited as evidence that elevated CO_2 increases soil C. In contrast, van Groenigen *et al*. (2006) is cited as evidence that responses are absent under low N conditions; de Graaff *et al*. (2006) is intermediate (Table 1). Thus, the conclusions derived from these meta-analyses are perceived to differ in the magnitude of the soil C response and how this response is influenced by N inputs.

Therefore, our first goal was to assess the influence of elevated CO_2 on soil C accumulation under low N conditions, examining the sensitivity of these estimates to different approaches used in the meta-analyses. For example, the four studies used different metrics to measure the effect of elevated CO_2 on soil C, and used different weighting functions to quantify the relative contribution of each study and how much influence each should have on the overall analysis. These four studies also used different criteria for including observations, extracted the data in different ways, and made different judgments about independence, all possibly influencing the magnitude of – and confidence in – the mean effect size estimate. How much influence did each of these factors have on the outcome of the meta-analyses, and thus on the inferences made about CO_2 effects on soil C?

Our second goal was to evaluate statistical power and biogeochemical significance. If low statistical power limits inferences drawn from individual experiments, it may also limit inferences drawn from a meta-analysis, even with the advantages of multiple studies. In the case of elevated CO_2 and soil C, finding no significant effect in the absence of exogenous N inputs could mean the effect is so small that it is not important. Alternatively, it could mean that estimates are highly variable, with a biogeochemically important result being indistinguishable from 'no effect.' To distinguish these possibilities, we compared estimated effects (and confidence intervals) with the residual terrestrial C sink (Schimel *et al*., 2001; Houghton, 2007) as a benchmark for biogeochemically significant changes in soil C caused by elevated CO_2.

Methods

Data compilation

We compiled data used in the four meta-analyses, including treatment means, sample sizes, variances, and experimental duration. We also constructed a composite dataset, comprising all observations included in any of the four individual meta-analyses. The composite data set included each effect size estimate from each of the four meta-analyses that represented an independent observation based on consensus criteria developed for this compilation. Independent observations included each possible elevated vs. ambient CO_2 treatment comparison in multifactor designs as long as the ambient CO_2 treatments were independent. In the case of multiple levels of elevated CO_2 and a single, ambient control, we used the average of the elevated CO_2 treatments vs. the single ambient CO_2 treatment, because multiple effect size estimates would rely on a single, nonindependent control, and because selecting a single CO_2 level would be arbitrary (because the multiple treatment levels were within the range of treatment CO_2 concentrations from other experiments with only a single elevated CO_2 treatment). We included experiments from all conditions, whether free air CO_2 enrichment (FACE), open top chambers (OTC), or greenhouse and growth chamber conditions. If only one meta-analysis included an experiment [e.g. Luo *et al*. (2006) was the only meta-analysis to include results from the salt marsh study (Drake *et al*., 1997)], that sole observation was included in the composite data set. In cases where more than one meta-analysis included an observation, we used the average of treatment means and variances in the composite data set, and then calculated effect sizes as described below. In some cases, the meta-analyses included multiple observations of effect sizes from a single study which by our criteria would not be considered independent, because the observations were all obtained from a single experimental treatment. These cases of nonindependence included multiple observations over time, from different soil depths, from different cover types, or from different soil fractions (e.g. mineral vs. organic). For meta-analyses that included such multiple nonindependent observations, we used the average of nonindependent effect size estimates (and variances) to represent that meta-analysis in the combined data set. For brevity, hereafter we refer to each data set only by the last name of the first author of the corresponding published paper (i.e. de Graaff, Jastrow, Luo, and van Groenigen).

Effect sizes

The four meta-analyses used different metrics for the effect of elevated CO_2 on soil C (Table 1). Even within each meta-analysis, often one effect size metric was selected for statistical purposes, but responses were converted to another to aid interpretation. The choice of an effect size metric can influence the outcome of a meta-analysis (e.g. Osenberg *et al*., 1997, 1999), so we compared how the results of each of the previous meta-analyses was affected by choosing different effect size metrics. We compared three different metrics of effect size:

Log Ratio. Jastrow and Luo used the log of the response ratio:

$$X_{LR} = \mathrm{Ln}(E/A), \tag{1}$$

where E is the mean soil C in the elevated CO_2 treatment and A is the mean soil C in the ambient CO_2 treatment. The log-ratio starts with an estimate of the relative change in C between the two treatment (E/A) and log-transforms it to improve its statistical behavior (Hedges *et al*., 1999). We note that this metric is perhaps best used to evaluate equilibrial changes, i.e. after trajectories following perturbations have restabilized (Osenberg *et al*., 1997, 1999).

Relative accumulation rate

Van Groenigen and de Graaff assumed that C accumulated linearly through time (and at a rate proportional to initial soil C concentrations). They calculated the effect of CO_2 as

$$CO_2\text{effect} = \ln(((E - A)/tA) + 1), \tag{2}$$

where E and A are as defined above and t is the duration of the experimental treatment (in years). Yet, to explain the biogeochemical significance of their findings, they used the relative accumulation rate

$$X_{RAR} = (EA/At)100\%. \tag{3}$$

This metric is more easily interpreted, and, over the range of effect sizes observed, it has nearly a linear relationship with Eqn (2). Thus, in this analysis we used the relative accumulation rate, with units of percent change per year. Note that this metric is related to Eqn (1) approximately by dividing by time [i.e. Eqn (3) is a rate; Eqn (1) is not], because $(E-A)/A$ approximates $\ln(E/A)$ for small differences between E and A.

Absolute accumulation rate

Although both Jastrow and Luo used the log of the response ratio to analyze findings statistically, they also both used the absolute rate of C accumulation to explain the biogeochemical significance of their results

$$X_{AAR} = (E - A)/t, \qquad (4)$$

where E and A are soil C pools in $g\,m^{-2}$ and t is the duration of the experiment, such that the metric has units of $g\,C\,m^{-2}\,yr^{-1}$. In the Jastrow meta-analysis, mean values of E and A were corrected for pretreatment soil C differences (when these data were available) by making additive adjustments to the average of all pretreatment soil C values before calculation of effect sizes (Jastrow *et al.*, 2005). Except where noted, these adjustments are included in the analyses of the Jastrow meta-analysis and apply to all effect size metrics.

Determining biogeochemical significance

Choosing an effect size metric is complex and open to debate, but we argue that it should estimate processes of interest and not be restricted to one or a few metrics that may be too easily applied to any scientific question (Osenberg *et al.*, 1997, 1999). In this case, a major process of interest is future soil C accumulation in response to rising atmospheric CO_2 concentrations. To place results from the meta-analyses into this context, we extrapolated absolute C accumulation calculated using Eqn (4) to the global scale. This allows comparing results from CO_2 experiments with current understanding of the size and cause of the modern residual terrestrial C sink [residual *sensu* Houghton (2007) because deforestation is not included in the estimate], estimated to be on the order of $2 \pm 1\,Pg\,C\,yr^{-1}$ (Schimel *et al.*, 2001; Fung *et al.*, 2005; Houghton, 2007). We sought to determine if our estimates of effects [using Eqn (4)] yielded effects on soil C that were large relative to this C sink. We therefore assumed that the short-term effects that we observed could be used to estimate the long-term response of soil C since preindustrial periods

$$S = X_{AAR}L([CO_2]_m - [CO_2]_p)/([CO_2]_e - [CO_2]_a), \qquad (5)$$

where S is the estimated change in soil C due to increased CO_2, L is global vegetated land area (excluding areas covered by sparse vegetation, permanent snow and ice, urbanization, and water) of 110 133 106 km^2 (Loveland *et al.*, 2000), X_{AAR} is the effect of elevated CO_2 on soil C as calculated by Eqn (4) (but expressed in $Pg\,C\,m^{-2}$), $[CO_2]_e$ and $[CO_2]_a$ are the elevated and ambient CO_2 levels in the experimental treatments, $[CO_2]_m$ and $[CO_2]_p$ are the modern (m) and preindustrial (p) CO_2 concentrations. Thus, the effect of X_{AAR} observed in going from $[CO_2]_e$ to $[CO_2]_a$ is assumed to apply proportionately to the historical change in going from $[CO_2]_p$ to $[CO_2]_m$.

This approach assumes that the dynamics of C accumulation caused by elevated CO_2 over 1–10 year step-change experiments can be extrapolated to gradual CO_2 rise over centuries, which is problematic for several reasons (Luo & Reynolds, 1999) and likely overestimates the actual rate of C accumulation in soils caused by elevated CO_2. On the other hand, the approach also assumes that effects of elevated CO_2 are a linear function of concentration (over the range considered, 280 to around 750 ppm), which underestimates the actual rate of C accumulation caused by elevated CO_2 to the extent this is a saturating function of concentration (e.g. Gill *et al.*, 2002). For purposes here, where our goal was not to independently estimate the size of the CO_2-driven soil C sink but rather to place results of the meta-analysis into a broader context, we feel that this approach is a reasonable starting point.

Weighting

We analyzed the individual data sets and the composite data set to test the sensitivity of the conclusions of the meta-analyses to the choice of weighting functions: weighting observations by the inverse of the pooled variance (Hedges & Olkin, 1985; Gurevitch & Hedges, 1999); weighting observations by a function of sample size where

$$\text{weight} = (N_a N_e)/(N_a + N_e), \qquad (6)$$

and N_a and N_e are the numbers of replicate observations for the ambient and elevated CO_2 treatments, respectively (Hedges & Olkin, 1985; Adams *et al.*, 1997); or weighting all observations equally. Weighting functions were used with all three of the effect size metrics defined above. In all cases, means and 95% confidence intervals were estimated using bootstrapping in META-WIN, a software package designed for meta-analysis (http://www.metawinsoft.com).

Nitrogen (N)

Three of the four meta-analyses examined the sensitivity of soil C accumulation to N inputs with elevated CO_2 (Luo, van Groenigen, de Graaff). We tested the sensitivity of the N effect to decisions about weighting and effect size metrics using the composite data set, the full combination of effect size metrics, and the three weighting functions described above. We also assessed the sensitivity of N fertilization effects to different inferential approaches, because the previous meta-analyses tested the effect of N on soil C accrual with elevated CO_2 in different ways. van Groenigen assigned all studies an N class, considering low N studies to have exogenous N supply of $<30\,\mathrm{kg\,ha^{-1}\,yr^{-1}}$, and high N studies to be above this amount. Luo and de Graaff included only experiments with high and low N treatments applied in a factorial design. We analyzed the composite data set using both approaches. In the first case, we assigned all studies to either 'low' or 'high' N, based on the cutoff of $30\,\mathrm{kg\,ha^{-1}\,yr^{-1}}$. For the second, we included only the subset of experiments where N treatments were crossed with CO_2 in a factorial design ($n = 15$). In this case, we used as our effect size estimate the difference in responses of soil C to elevated CO_2 observed between high and low N treatments.

Differences

Even if all four meta-analyses had used the same metric and weighting scheme for calculating the effect size of elevated CO_2 on soil C, the values of those average effect sizes would differ for at least three reasons: (1) Different 'Studies': different studies were included in the meta-analyses; (2) 'Extraction': different estimates of the effect size were extracted from the same studies, for example, because different data sources were used for the same studies or because of errors during figure scanning. (For expediency, this category also includes effect size differences resulting from Jastrow's adjustments for pretreatment soil C differences described above in 'Effect sizes,' although these cases were omitted to isolate components of 'extraction,' as described below. Overall, excluding observations adjusted for pretreatment differences in soil C caused less than a 30% change in the influence of 'extraction,' so the pretreatment adjustment does not appear to dominate.); and (3) 'Independence': data from experiments were summarized using different criteria involving decisions about independence. We evaluated the relative influence of each of these using the three effect size metrics. We quantified components of the difference between mean effect size estimates for the different meta-analyses, decomposing these differences into the three components described above ('Studies,' 'Extraction,' and 'Independence').

The logic describing the derivation of these three components follows. Consider two arrays, X and Y, and here representing arrays of effect size estimates from two meta-analyses. The difference (D) between the means of X and Y can be written:

$$D = \Sigma X_i/n_X - \Sigma Y_i/n_Y, \tag{7}$$

where ΣX_i and ΣY_i are the sums of elements in arrays X and Y and n_x and n_y are the total numbers of elements in each array.

X and Y may share elements in common (C), but they may differ in value (V_X and V_Y) due to extraction errors. Further, some elements may be unique to each array. The uniqueness arises for two reasons: (1) one meta-analysis included data from a study not included in the other meta-analysis ('Studies'); and (2) one meta-analysis used more effect sizes from the same study because they used less stringent criteria about independence ('Independence'). Thus, there are four ways to describe the possible elements in the two arrays:

- Common and identical in value (C).
- Common but different in value due to extraction errors (V).
- Unique because a different study was used (S).
- Unique because more lenient criteria for independence was used (I).

Eqn (7) can be expanded by subsetting the total elements into these four components

$$\begin{aligned} D = {} & (\Sigma C + \Sigma V_x + \Sigma S_x + \Sigma I_x)/(n_C + n_V + n_{X,S} \\ & + n_{X,I} - (\Sigma C + \Sigma V_y + \Sigma S_y + \Sigma I_y)/(n_C + n_V \\ & + n_{Y,S} + n_{Y,I}, \end{aligned} \tag{8}$$

where ΣC is the sum of identically valued elements common to both arrays, ΣV_X and ΣV_Y are the sums of common elements that differ in value between X and Y (e.g. different values were extracted from the same published figure), ΣS_X and ΣS_Y are the sums of elements unique to each array because they are derived from different studies, ΣI_X and ΣI_Y are the sums of elements unique to each array because one meta-analysis used more lenient criteria about independence (i.e. more effect sizes were included from one study in one of the meta-analyses), and n gives the number of effect sizes in each subset.

Rearrangement yields

$$\begin{aligned} D = {} & [(n_C + n_V + n_{y,S} + n_{Y,I})(\Sigma C + \Sigma V_x + \Sigma S_x + \Sigma I_x) \\ & - (n_C + n_V + n_{x,S} + n_{x,I})(\Sigma C + \Sigma V_y + \Sigma S_y \\ & + \Sigma I_x)]/(n_y n_x), \end{aligned}$$

(9) which simplifies to,

$$\begin{aligned} D = {} & [(n_{y,S} + n_{y,I})\Sigma C - (n_{x,S} + n_{x,I})\Sigma C + n_y \Sigma I_x \\ & - n_x \Sigma I_y + n_y \Sigma S_x - n_x \Sigma S_y - n_y \Sigma V_x \\ & - n_x \Sigma V_y]/(n_y n_x). \end{aligned} \tag{10}$$

Elements that are identical in value and shared between the two arrays (ΣC) do not cause differences between them. Rather, changing the numbers of elements in each array (n_x and n_y) affects the relative influence of common elements on the grand means, and thus their difference (D). We therefore partitioned and ascribed this influence into those factors that actually alter n: including different studies (S) and using more lenient criteria for independence (I). We then decomposed Eqn (10) into three components:

$$D_{\text{Independence}} = \{(n_y \Sigma I_x - n_x \Sigma I_y) + \Sigma C(n_{y,I} - n_{x,I})\}/(n_x n_y), \quad (11)$$

$$D_{\text{Studies}} = \{(n_y \Sigma S_x - n_x \Sigma S_y) + \Sigma C(n_{y,S} - n_{x,S})\}/(n_x n_y), \quad (12)$$

$$D_{\text{Extraction}} = \{(n_y \Sigma V_x - n_x \Sigma V_y)/(n_x n_y). \quad (13)$$

We used these equations to quantify how much each of these components contributed to the differences among meta-analyses in mean estimates of the effect size.

Components of extraction

We further explored the differences in effect size estimates associated with data extraction. In most cases, meta-analyses used the same data source to estimate effect sizes from a given experimental system, often because only one data source was available. In other cases, though, more than one publication included estimates of soil C responses to elevated CO_2 from a single experimental system, and different meta-analyses used different published papers to estimate effect sizes. We assessed all paired comparisons of meta-analyses where the same data source was used, and where different data sources were used, in each case calculating the absolute value of the difference in effect size estimates. In all these comparisons, we excluded cases where independence had been treated differently and where soil C contents had been adjusted for pretreatment differences.

Relaxed vs. strict approach to independence

We further explored the influence of independence by comparing the effect of including multiple observations vs. deriving a single estimate from each experimental comparison. We constructed a data set including all experiments in which approaches to independence differed among the four meta-analyses (Table 3) to compare 'relaxed' vs. 'strict' approaches to statistical independence. In the relaxed case, we used all effect size observations from each experiment, derived from multiple observations over time, over soil depths, or stratified samples all within a single experimental comparison (i.e. multiple observations derived from a single set of elevated and ambient plots). In the strict case, we included multiple effect size estimates from multifactor experiments because there are independent controls for each treatment. However, a single factor design with many levels would yield only one effect size estimate, because all treatments are compared with the same control. We used METAWIN to calculate the mean effect sizes and the bootstrapped 95% confidence intervals for the low and high N conditions under the four weighting schemes described previously.

Criteria for inclusion

The four meta-analyses also differed in criteria for study inclusion, a common practice in meta-analysis (Englund *et al*., 1999). We therefore explored the consequences of excluding studies based on assessments of quality or relevance. In this case, we compared responses across the entire data set with responses from experiments that arguably best represent responses of the nonagricultural terrestrial surface (i.e. most of it): field experiments that used open-top chamber or free-air CO_2 enrichment technologies, that were conducted in undisturbed (untilled) soils, that lasted 2 years or longer, and that occurred without exogenous N inputs. The first three of these criteria roughly parallel those used by Jastrow to construct the data for meta-analysis, and the fourth was a component of the meta-analyses of de Graaff, Luo, and van Groenigen. Together, these criteria capture common perceptions of what constitutes global change field experiments that are most relevant to understanding future responses of the Earth's nonagricultural terrestrial ecosystems.

Results

Metrics and weights

For each metric, there was a range in the estimates of the mean effect size across the 12 combinations of three weighting schemes and four meta-analyses under low N conditions (Fig. 1, Table 2). For example, the percent change in soil C per year estimates ranged from −0.03% to 1.36% yr^{-1} (Table 2), and estimates of the CO_2 effect on soil C accumulation ranged from −6.5 to 26.9 $g\,m^{-2}\,yr^{-1}$ (Fig. 1). These differences are biogeochemically important. Expressing accumulation as a global flux driven by the observed increase in atmospheric CO_2 [using Eqn (5)], they correspond to C sinks of −0.3 to 1.1 $Pg\,C\,yr^{-1}$ (Table 2), from a small source to around half of the current residual sink (Houghton, 2007).

Within each of the data sets from the four separate studies, the use of different weighting functions had some influence on the mean estimates of the effect size (Fig. 1, Table 2), although the influence of the weighting function within one meta-analysis was smaller than the differences among meta-analyses. Within one meta-analysis, weighting by sample size or with no weight gave comparable estimates that were often larger than weighting by the inverse of the variance (Fig. 1, Table 2). Despite the differences in the mean effect size estimates across data sets and weighting functions, the 95% confidence intervals of all 12 estimates overlapped (Table 2, and grey band in Fig. 1). Therefore, although the differences among means were large and biogeochemically meaningful, confidence in these differences was limited.

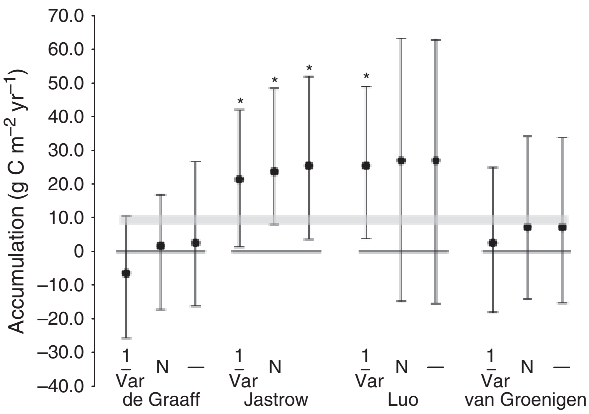

Fig. 1 The influence of different weighting schemes on the estimate of the effect of elevated CO_2 on soil C accumulation (g C m^{-2} yr^{-1}) at low N supply according to the four meta-analyses, identified by the lead author's surname. Filled circles are means and vertical bars 95% confidence intervals (*confidence intervals that do not overlap zero). Weighting functions are the inverse of the variance in the effect size estimate (1/Var), a function of sample size (*N*, Adams *et al.*, 1997), and no weight (–). Gray shaded area indicates region of overlap for all 12 confidence intervals. Dashed horizontal line indicates an effect of zero.

Even though the 95% confidence intervals all overlapped with each other, they did not all overlap zero, arguably leading to different inferences about the presence or absence of an effect of elevated CO_2 at low N when based on null hypothesis tests (see Osenberg *et al.*, 2002). Specifically, the 95% confidence intervals overlapped with zero for all cases in the van Groenigen and de Graaff data sets, but never did in the Jastrow data set; the Luo data set was intermediate (two of three confidence intervals overlapped zero). In all cases, when confidence intervals did not overlap zero, they indicated a positive effect of elevated CO_2 on soil C accumulation under low N conditions. Thus, the different inferences drawn from these four meta-analyses cannot be explained entirely by differences in effect size definitions or weights.

Studies included, data extraction, and independence

The four meta-analyses had only modest overlap among the CO_2 experiments they included in their

Table 2 Comparison of databases from four published meta-analyses for the effect of elevated carbon dioxide (CO_2) on soil carbon (C) under low nitrogen conditions, and the influence of the effect size metric and weighting function

		X_{LR}		X_{RAR}		S	
Database	Weighting	Mean	95% CI	Mean (%)	95% CI	Mean	95% CI
de Graaff	1/Var	−0.004	−0.045 to 0.033	−0.03	−1.99 to 1.88	−0.28	−1.12 to 0.44
	N	−0.009	−0.049 to 0.030	0.26	−1.58 to 2.10	0.31	−0.67 to 1.43
	–	−0.003	−0.049 to 0.037	0.26	−1.59 to 2.14	0.31	−0.62 to 1.45
Jastrow	1/Var	0.042	0.014 to 0.066	0.82	0.20 to 1.41	0.95	0.05 to 1.86
	N	0.030	0.007 to 0.057	0.91	−0.23 to 2.15	1.13	0.15 to 2.30
	–	0.055	−0.003 to 0.075	0.92	−0.23 to 2.20	1.13	0.16 to 2.16
Luo	1/Var	0.047	0.013 to 0.080	1.35	0.54 to 2.22	1.01	0.33 to 2.05
	N	0.061	0.026 to 0.101	1.36	0.50 to 2.16	1.15	−0.64 to 2.70
	–	0.059	0.016 to 0.083	1.35	0.51 to 2.09	1.15	−0.67 to 2.67
van Groenigen	1/Var	0.015	−0.011 to 0.039	0.01	−1.53 to 1.50	0.07	−0.73 to 0.67
	N	−0.007	−0.042 to 0.022	0.13	−1.33 to 1.60	0.10	−0.68 to 1.08
	–	−0.004	−0.035 to 0.026	0.15	−1.40 to 1.65	0.10	−0.76 to 1.01

Weighting functions assessed include weighting by the inverse of the pooled variance (1/Var), weighting by the number of experimental replicates (Adams *et al.*, 1997), and no weighting (–). The effect size metrics include X_{LR}, the log of the response ratio, X_{RAR} or the relative accumulation rate expressed as the percent change in soil C per year of experimental CO_2 exposure, and S, which is the extrapolation of the observed CO_2 effect on soil C accumulation to the global scale, here expressed in units of Pg C yr^{-1} [using Eqn (5)]. (X_{AAR} is shown in Fig. 1). The values in parentheses are the bootstrapped lower and upper 95% confidence intervals.

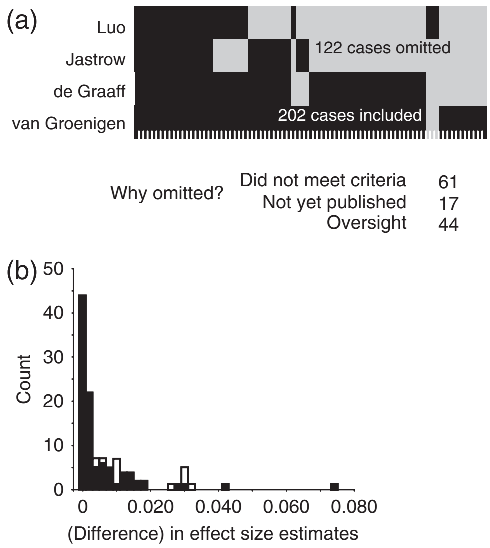

Fig. 2 (a) Overlap (and lack thereof) in the 81 observations included in at least one of the four meta-analyses. For each of the four meta-analyses (identified by last name of first author on the left), black shading indicates cases that were included, gray indicates cases that were omitted. White hatch marks on the bottom indicate the scale (one mark, one case). The text within the box indicates the total number of cases included by the four meta-analyses (202) and the total number of cases that were omitted (122). The table underneath shows how many studies were excluded because they did not meet the authors' criteria, omitted because the data had not yet been published, or omitted because of oversight. (b) Frequency distribution of the absolute value of the difference in effect size estimates between individual meta-analyses for the metric, percent change per year. For a given experiment, we calculated the difference in effect size estimates between each pair of meta-analyses that assessed that experiment. Observations were excluded in cases where independence was treated differently or where pretreatment adjustments of soil C content had occurred, for a total of 114 unique pairs. To isolate further the components of 'extraction' errors, the figure shows cases where identical data sources were used to estimate effect sizes (filled bars), and cases where different data sources were used (open bars).

meta-analyses (Fig. 2a). The lack of overlap among studies partly reflects differences among the four meta-analyses in the criteria they used for study selection. For example, whether to include studies conducted in greenhouses and growth chambers (e.g. van Groenigen, Luo), or only 'field' studies (de Graaff, Jastrow), and whether to exclude studies lasting less than one (van Groenigen) or two growing seasons (Jastrow). Our survey of the lead authors showed that failure to meet stated criteria explained most omissions, and a handful of studies were omitted because they were published too late to be included in a given meta-analysis. Yet, a surprising number of cases were omitted because the search algorithms used to gather appropriate literature were not exhaustive (Fig. 2a).

The four meta-analyses also differed in data extraction. Data compiled by the four meta-analyses usually did not yield the same estimates of the effect size of elevated CO_2 on soil C for the same study, even after removing cases where independence was treated differently and cases where values of soil C had been adjusted for pretreatment differences (Jastrow *et al*., 2005), and after converting all data to the same effect size metric (Fig. 2b). Of 114 paired comparisons of effect size estimates for the same studies among the four meta-analyses, in only 44 cases did two meta-analyses come up with identical estimates. When meta-analyses used the same data source (99 cases), the average difference in effect size estimates was smaller ($0.43 \pm 0.10\%\ \mathrm{yr}^{-1}$, mean ± standard error) than when meta-analyses used different data sources (15 cases, mean difference $1.55 \pm 0.30\%\ \mathrm{yr}^{-1}$). Thus, using data from separate samplings of a given experiment [e.g. Williams *et al*. (2000) vs. Jastrow *et al*. (2005) for the tallgrass prairie study; or Johnson *et al*. (2004) vs. Jastrow *et al*. (2005) for the Sweetgum forest study] caused larger differences in effect size estimates than did relying on the same data source. Nevertheless, whether using the same data source or different data sources, the mean differences in effect size estimates (0.43% and $1.55\%\ \mathrm{yr}^{-1}$) were rather large, comparable with some estimates of the mean effect size itself (see X_{RAR} in Table 2).

The four meta-analyses also differed in what were considered independent estimates of effect size (Table 3). The meta-analyses consistently included multiple estimates from multifactor experiments (34 cases), and single estimates from single-factor experiments (i.e. manipulating CO_2 only, 12 cases). However, in a number of cases, the meta-analyses took different approaches toward what constitutes statistically independent observations (Table 3).

The observed differences in effect sizes under low N conditions were influenced by each of the three factors we investigated, but of the three, the decision about what constituted an independent sample had the largest influence (Table 4). Calculated as a component of the overall magnitude of the difference in effect sizes between paired meta-analyses, approaches to independence caused 45–70% of the difference between effect size estimates, whereas the inclusion of different studies caused 5–24%, and data extraction contributed 19–31%. Calculated on a per case basis, in all but one case (van Groenigen–de Graaff, for 'accumulation'), independence had the largest influence for all metrics considered (Table 4). This indicates that decisions about what

Table 3 Number of additional observations included from experiments in which independence was treated differently by the four different meta-analyses across the entire dataset, showing the specific experiment from which multiple observations were drawn, the 'source' of those multiple observations, and the additional number of observations (' + df', or the increase in degrees of freedom) that resulted from including multiple observations from a single experiment in those meta-analyses that used more relaxed criteria for independence. The footnotes show which meta-analyses used the more lenient criteria in particular cases.

Experiment	Source	+ df	References
Desert FACE	Stratified sampling by cover type*	2	Billings *et al*. (2002)
Loblolly pine FACE	Stratified sampling by depth†, multiple soil fractions†, and multiple samples in time†	7	Lichter *et al*. (2005), Schlesinger & Lichter (2001)
Chaparral	Multiple CO_2 treatments‡	2	Treseder *et al*. (2003)
Ponderosa Pine	Multiple CO_2 treatments‡ and multiple samples in time†	8	Johnson *et al*. (1997, 2000)
PopFACE	Nested plots*,§	2	Hoosbeek *et al*. (2004)
Swiss FACE	Multiple samples in time†	4	van Kessel *et al*. (2006), Six *et al*. (2001), de Graaff *et al*. (2004), van Groenigen *et al*. (2002), van Kessel *et al*. (2000)
Crops	Multiple soil fractions†	2	Prior *et al*. (2004), Torbert *et al*. (2000)

*de Graaff.
†Luo.
‡Jastrow.
§van Groenigen.

constitute independent samples in meta-analyses can strongly influence the meta-analyses' results.

Combined data set

Using the combined data set, which represents the 'average' across all four meta-analyses under all N conditions, we found that elevated CO_2 significantly increased soil C, regardless of metric or weighting function (Table 5). Thus, when averaged across all studies and N conditions, these meta-analyses support a similar conclusion about the overall effect of elevated CO_2. We also found that added N caused a larger increase in soil C accumulation with elevated CO_2, regardless of approach, metric or weighting scheme (Table 5). Including only experiments in which N and CO_2 were manipulated in a factorial design, the difference between CO_2 effects at high and low N was positive, indicating that CO_2 had larger effects at high N. The 95% confidence intervals did not always exclude zero, reflecting the lower sample size for the factorial design, but the N effects were comparable in magnitude and direction to those estimated from including all observations (Table 5).

In all cases, the effect of elevated CO_2 at low N supply was not significantly different from zero, though the confidence intervals included biogeochemically significant rates of soil C accumulation. For example, the confidence intervals for the effect of elevated CO_2 on soil C accumulation at low N ranged from -10 to $+39\,g\,m^{-2}\,yr^{-1}$, with the upper estimate potentially explaining a $1.6\,Pg\,C\,yr^{-1}$ residual terrestrial sink [calculated using Eqn (5)]. Thus, while meta-analysis can show the influence of added N on the CO_2 effect, the data currently available provide limited power to address the significance of the effect of elevated CO_2 on soil C in the absence of N addition.

Including multiple observations from single experiments ('relaxed' independence) caused higher estimates of the effect of elevated CO_2 on soil C under low N conditions, and lower estimates under high N conditions, compared with the stricter approach of including only a single observation from each experimental comparison ('strict' independence, Fig. 3). This pattern held for all three weighting functions considered. Including multiple estimates also narrowed the confidence intervals, under low N conditions, resulting in confidence intervals that did not overlap with zero (Fig. 3).

Restricting the data set to only those field studies that used open-top chambers or free-air CO_2 enrichment, that occurred in undisturbed soils, and that lasted at least 2 years did not alter the effect of elevated CO_2 on soil C (Table 5). Estimates of mean soil C accumulation rates from these more 'realistic' experiments were similar to those from the full combined data set under low N conditions, which included short duration and growth chamber experiments; in both cases, confidence intervals all overlapped zero, indicating no statistically significant effect of elevated CO_2. As with the full combined data set under low N conditions, confidence

Table 4 The relative influence of independence, studies included, and data extraction on the differences between mean effect size estimates among the four meta-analyses under low N conditions for each of the three metrics of effect size considered, X_{LR} ($\times$ 100 for easier comprehension), X_{RAR}, and X_{AAR}

Metric	Luo–de Graaff	Luo–Jastrow	Luo–van Groenigen	Jastrow–de Graaff	Jastrow–van Groenigen	van Groenigen–de Graaff
Log response ratio (X_{LR}) $\times$ 100						
Overall magnitude						
Independence	3.8	3.3	3.9	1.4	2.0	−0.7
Studies	1.9	−1.2	1.6	1.5	0.2	0.9
Extraction	0.7	0.3	0.8	1.2	1.6	0.0
Per case influence						
Independence	0.27	0.22	0.28	0.20	0.18	0.18
Studies	0.07	0.08	0.04	0.07	0.01	0.07
Extraction	0.06	0.04	0.05	0.06	0.07	0.00
Relative accumulation rate (X_{RAR})						
Overall magnitude						
Independence (%)	0.79	0.68	0.79	0.40	0.66	−0.25
Studies (%)	0.22	−0.33	0.19	0.14	−0.21	0.23
Extraction (%)	0.19	0.09	0.22	0.22	0.33	0.01
Per case influence						
Independence (%)	0.06	0.05	0.06	0.06	0.06	0.06
Studies (%)	0.01	0.02	0.01	0.01	0.01	0.02
Extraction (%)	0.02	0.01	0.02	0.01	0.01	0.00
Absolute accumulation rate (X_{AAR}) ($g\,m^{-2}\,yr^{-1}$)						
Overall magnitude						
Independence	8.6	4.3	8.7	5.4	11.8	−0.2
Studies	9.9	−1.2	9.7	5.5	−0.9	−2.3
Extraction	2.6	−1.5	6.0	8.4	11.8	−0.7
Per case influence						
Independence	0.6	0.3	0.6	0.8	1.1	0.1
Studies	0.3	0.1	0.3	0.2	0.0	0.2
Extraction	0.2	0.3	0.4	0.4	0.5	0.1

For each metric, we calculated the overall magnitude of the influence of independence, studies included, and data extraction [Eqns (11–13)]; the sum of these three values is the difference between the two means. We also calculated the per case influence, which is the absolute value of the overall magnitude, divided by the frequency of cases (i.e. observations) where independence was treated differently, different studies were included, or data extraction resulted in different effect size estimates, for each comparison.

intervals from the 'realistic' experiments encompassed a range that included biogeochemically meaningful effect sizes; e.g. the upper 95% confidence limit for the estimate of accumulation weighted by sample size was $56.9\,g\,C\,m^{-2}\,yr^{-1}$, corresponding to a residual sink of $2.5\,Pg\,C\,yr^{-1}$. Thus, selecting experiments representing arguably more realistic conditions had little influence on the outcome of the meta-analysis.

Discussion

The range in mean estimates and confidence intervals resulting solely from using different weights (Fig. 1, Table 2) illustrates the potential for the selection of a weighting function alone to affect inferences drawn from meta-analysis, inferences about both the magnitude of the effect size and its significance. Weighting observations is standard procedure in meta-analysis (Gurevitch & Hedges, 1999), but assessing the influence of different weighting functions is not, perhaps because the use of default weighting functions (like the inverse of the pooled variance) is perceived to be a prescribed part of sound meta-analytic procedure (Hedges & Olkin, 1985). For one data set (Luo), selection of a different weighting function yielded a statistically significant result in one case and nonsignificant results in two other cases. Our findings therefore underscore the importance of justifying the selection of any chosen weighting function and assessing the influence of alternative selections on the outcome of the meta-analysis.

Including different studies, extracting data, estimating different effect sizes, and approaching indepen-

Table 5 Summary of meta-analysis results for the combined dataset, for three different metrics and three different weighting schemes (defined in 'Methods')

		The whole dataset		Low N		High N		N effect	Factorial experiments only		Field experiments, at least 2 years, low N	
Metric	Weighting	Effect size	95% CI	Effect size	95% CI	Effect size	95% CI	*P*-value	High N– Low N	95% CI	Effect size	95% CI
X_{LR}	1/Var	4.3	1.8 to 6.6	1.7	−1.2 to 4.7	7.6	3.6 to 11.3	0.022	7.4	0.6 to 1.3	2.2	−3.6 to 6.4
(×100)	*N*	3.1	0.7 to 5.3	1.0	−2.1 to 4.3	6.2	3.1 to 9.5	0.046	3.7	−0.9 to 8.5	1.6	−2.4 to 5.6
	–	3.4	1.4 to 5.7	0.4	−02.7 to 3.1	7.2	3.9 to 10.7	0.002	4.8	−0.3 to 9.5	0.4	−3.4 to 3.7
X_{RAR}	1/Var	1.30%	0.36 to 2.23	0.23%	−1.19 to 1.40	2.55%	1.23 to 4.08	0.010	1.9%	0.3 to 4.2	0.2%	−0.9 to 1.3
	N	1.45%	0.32 to 2.52	0.34%	−0.97 to 1.66	2.83%	1.43 to 4.41	0.016	1.7%	−0.2 to 4.0	0.6%	−0.5 to 1.6
	–	1.48%	0.42 to 2.53	0.35%	−0.96 to 1.79	2.87%	1.41 to 4.56	0.029	2.2%	−0.2 to 4.6	0.2%	−0.9 to 1.2
X_{AAR}	1/Var	18.8	5.7 to 33.6	13.0	−8.2 to 39.4	25.1	12.0 to 40.4	0.420	28.5	−4.7 to 57.2	3.1	−10.8 to 56.1
	N	26.7	8.0 to 49.8	9.2	−10.0 to 30.3	48.4	17.8 to 90.0	0.054	60.9	4.2 to 154.0	23.9	−3.9 to 56.9
	–	26.7	8.2 to 49.2	9.2	−10.4 to 32.8	48.4	18.7 to 91.9	0.058	68.4	8.3 to 161.8	12.4	−9.5 to 36.4

Shown are the grand mean effect sizes across all N treatments for the whole data set, with the corresponding 95% confidence interval. 'Low N' and 'High N' show the mean effect sizes for the different N categories, with corresponding 95% confidence intervals. '*P*-value, N effect' is the probability that the low and high N categories are equal. 'Factorial experiments only' shows the mean difference and 95% confidence interval in response to CO_2 caused by N addition (high–low N) based on experiments where CO_2 and N were crossed in a factorial design ($n = 15$ experiments). The last two columns show mean effect sizes and 95% confidence intervals for an analysis excluding studies that lasted <2 years, were conducted in disturbed soils, greenhouses or glasshouses, or with exogenous N additions $>30\,\mathrm{kg\,N\,m^{-2}\,yr^{-1}}$ ($n = 25$ observations).

dence differently all affected the outcome of the meta-analysis, but independence had the largest influence (Table 4). Independence may have had a larger effect than the other factors considered because repeated measurements of a similar effect size will tend to put greater emphasis on that particular study (and its effect) Criteria for what constitutes an independent sample suitable for inclusion in meta-analysis are neither clear nor uniformly applied. Including multiple effect size estimates from all control vs. treatment comparisons in a multifactor experiment is often considered a reasonable balance (e.g. Gurevitch *et al*., 1992; Curtis & Wang, 1998; Bancroft *et al*., 2007; Morris *et al*., 2007), even though those comparisons are arguably not independent because they occur within the same experimental setting. Some meta-analyses include multiple observations from the same experimental comparison – for example, including multiple observations over time from studies of leaf decomposition rates (Knorr *et al*., 2005), or multiple observations from different soil depths in elevated CO_2 experiments (Luo *et al*., 2006). The argument for this approach is that excluding or even averaging across multiple observations within a given study has too high a cost in lost information. One argument for stricter approaches to independence is that it reduces the chance of making flawed inferences (Hedges & Olkin, 1985). Another important argument is that not all times (or levels) may be relevant to the question being addressed, so data should be restricted to the timescale most relevant to the question and effect size metric (Osenberg *et al*., 1997, 1999; Downing *et al*., 1999).

Our comparison shows that the approach to independence can influence the outcome of the meta-analysis: using the 'strict' approach provides weak evidence of a significant CO_2 effect on soil C accrual at low N, whereas the 'relaxed' approach usually indicates that the effect is significant (Fig. 3). One possibility is to avoid including multiple nonindependent observations and include the information in some other way, for example assigning higher weights to means of multiple observations, or by restricting analysis to the most appropriate data for the question (Downing *et al*., 1999). Minimally, we suggest that if practitioners choose to include multiple nonindependent observations, they should assess the impact of this decision on the outcome of the meta-analysis, and interpret the findings with caution if one approach indicates a 'significant' effect but the other does not.

Although it is not surprising that including or excluding particular studies influenced the mean effect size estimate (Table 4), it was surprising how little the meta-analyses overlapped in the studies they included. There were three reasons for the lack of overlap. First, the time periods over which data were gathered were not identical, so some meta-analyses had access to more recent data than others. Second, the ability to find appropriate studies was also important: clearly, the search methods

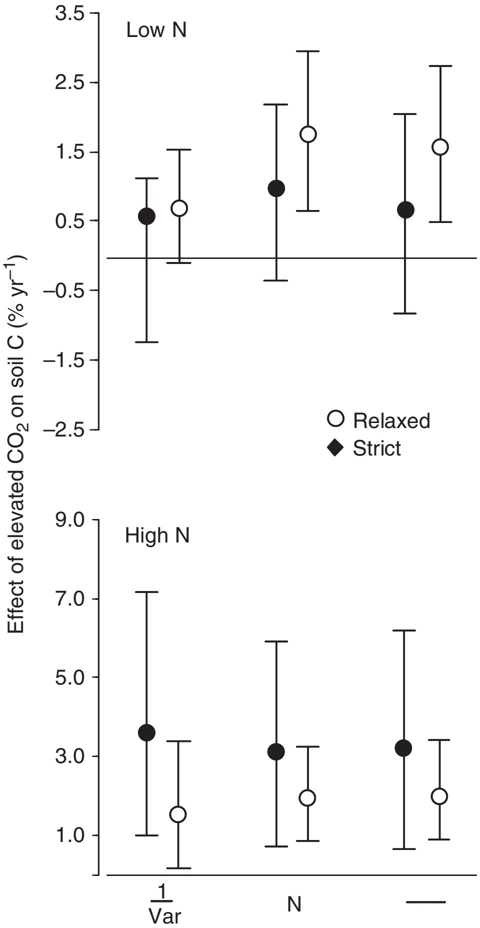

Fig. 3 The effect of different approaches to independence on mean effect size (measured as percent per year) and 95% confidence intervals for the effect of elevated CO_2 on soil C. Data were compiled from all cases where the four meta-analyses used different approaches to independence ($n = 20$ for 'strict', 55 for 'relaxed'). Weighting functions are indicated below the horizontal axis: – (no weight), 1/Var (inverse of the variance), and *N* [Eqn (6)]. For definitions of 'strict' and 'relaxed' independence, see 'Methods'.

used were not always exhaustive, such that studies that had been published and would have met the relevant criteria were simply missed. Good practice may thus require using multiple search tools, and requesting that reviewers of manuscripts reporting meta-analysis evaluate the completeness of the search. More broadly, this points to the need to overcome a long-standing problem in data synthesis in ecology: poor adoption and availability of public data archival systems (Parr & Cummings, 2005). Third, the meta-analyses used different criteria for what constituted an acceptable study. Evaluating study quality is common in meta-analysis and is entirely reasonable. In the present case, long-term experiments conducted in the field with intact soils are probably more appropriate for assessing the likely future role of terrestrial soils for sequestering C than are short-term experiments in greenhouses using disturbed soils mixed with sand. But even with such defensible criteria for study exclusion, an alternative approach is to include all the data and test empirically whether selection criteria influence the outcome. If so, the criteria for exclusion may be justified. If not (as in the present case), the test with the broader data set will be more powerful (Englund *et al.*, 1999).

Our finding that the response of soil C depended on exogenous N supply is not surprising given that the three meta-analyses which separately assessed the effect of N also found this (de Graaff *et al.*, 2006; Luo *et al.*, 2006, van Groenigen *et al.*, 2006). The consistency of the response across effect size metrics, weighting functions, and inference tests (i.e. whether across the whole data set or restricted to factorial experiments) further suggests this result is robust. What drives this N response? Past syntheses show that N addition usually increases soil C accumulation, a conclusion supported by meta-analyses in forest ecosystems (Johnson & Curtis, 2001), experiments in grasslands (Dijkstra *et al.*, 2005; Billings *et al.*, 2006, but see Xie *et al.*, 2005), and meta-analyses in crops (Alvarez, 2005). For crops, increased soil C required returning crop residues to the soil (Alvarez, 2005), consistent with the notion that the positive effects of N addition on soil C accumulation are an indirect response to increased plant growth with N addition. The direct effects of N addition on organic matter decomposition can be positive, neutral, or negative (Neff *et al.*, 2002), and in some cases the negative effects dominate the response at the ecosystem level, even when added N increases plant production (e.g. Mack *et al.*, 2004). Thus, the enhanced response of soil C to elevated CO_2 with added N could reflect both the synergistic effect of N and CO_2 on plant production and interactions between N and CO_2 on decomposition of soil organic matter (Reich *et al.*, 2006).

Meta-analysis in ecology has been lauded as a powerful synthetic tool (Gurevitch & Hedges, 1999; Osenberg *et al.*, 1999), and use of meta-analysis has increased dramatically in the past decade (Fig. 4). With meta-analysis, idiosyncratic results from a handful of studies can be appropriately categorized as odd, nongeneralizable, and unimportant for explaining broad-scale patterns. On the other hand, a collection of nonsignificant results, when considered together, can reveal general patterns otherwise inscrutable against the backdrop of natural variation, as is often the case in individual studies examining soil C (Conen *et al.*, 2003). As with any statistical technique, meta-analysis is sensitive to subjective decisions made during its application. We also acknowledge that interpretations of meta-analyses, like any data interpretation, have an element of sub-

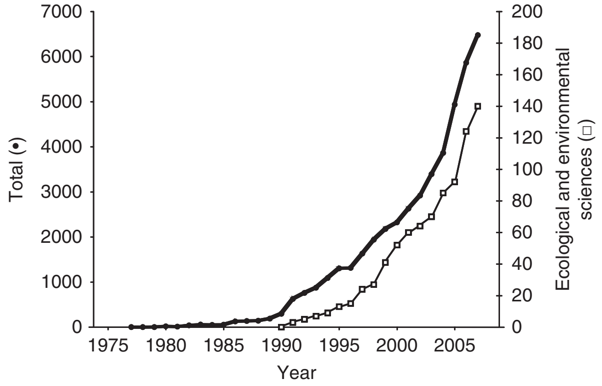

Fig. 4 Increasing use of meta-analysis in the published scientific literature. The following databases were surveyed for occurrence of the term 'meta-analysis': BIOSIS, Engineering Index, Inspec, ISI Art & Humanities, ISI Social SciSearch, and ISI SciSearch, using the SEARCHPLUS tool version 2.4 (left axis, filled circles). Searches were repeated, but restricted to specific categories (ecology, environmental sciences, fisheries, biodiversity conservation, marine & freshwater biology, soil science, water resources) relevant to the ecological and environmental sciences (right axis, open squares).

jectivity, and the choice of authors to emphasize one point over another can influence how those papers are perceived.

Results from meta-analyses can change with different criteria for data selection (Englund *et al*., 1999), with the metric of the effect size selected (Osenberg *et al*., 1997, 1999), and as shown here, with data extraction, weighting functions, study selection, and approaches to statistical independence. We encourage researchers applying meta-analysis to assess and report the sensitivity of their findings to such decisions, and to clarify their procedures. Specifically, we recommend that meta-analysts: (1) state specifically over what time period studies were gathered, particularly the point at which admission of new studies ceased; (2) describe search algorithms, including search terms; (3) take an exhaustive approach to data gathering and test empirically whether assessments of study 'quality' influence the outcome; (4) use metrics of effect size that reflect the specific question of interest; and (5) conduct multiple meta-analyses with the same data set, examining whether weighting functions, effect size metrics, and decisions about independence influence the outcome. Results that are highly sensitive to subjective decisions should be viewed with caution.

Acknowledgements

Thanks to participants in the Merriam-Powell Seminar Series at Northern Arizona University for useful feedback on the design and analysis. This research was supported by the National Science Foundation (DEB-0092642 and DEB-0543218) and the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division under contract DE-AC02-06CH11357.

References

Adams DC, Gurevitch J, Rosenberg MS (1997) Resampling tests for meta-analysis of ecological data. *Ecology*, **78**, 1277–1283.

Alvarez R (2005) A review of nitrogen fertilizer and conservation tillage effects on soil organic C storage. *Soil Use and Management*, **21**, 38–52.

Bancroft BA, Baker NJ, Blaustein AR (2007) Effects of UVB radiation on marine and freshwater organisms: a synthesis through meta-analysis. *Ecology Letters*, **10**, 332–345.

Billings SA, Brewer CM, Foster BL (2006) Incorporation of plant residues into soil organic matter fractions with grassland management practices in the North American Midwest. *Ecosystems*, **9**, 805–815.

Billings SA, Schaeffer SM, Zitzer S, Charlet T, Smith SD, Evans RD (2002) Alterations of nitrogen dynamics under elevated CO_2 in an intact Mojave Desert ecosystem: evidence from δ^{15}N. *Oecologia*, **131**, 462–467.

Conen F, Yakutin MV, Sambuu AD (2003) Potential for detecting changes in soil organic C concentrations resulting from climate change. *Global Change Biology*, **9**, 1515–1520.

Cramer W, Bondeau A, Woodward FI *et al*. (2001) Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373.

Curtis PS, Wang XZ (1998) A meta-analysis of elevated CO_2 effects on woody plant mass, form, and physiology. *Oecologia*, **113**, 299–313.

De Graaff MA, Six J, Harris D, Blum H, van Kessel C (2004) Decomposition of soil and plant C from pasture systems after 9 years of exposure to elevated CO_2: impact on C cycling and modeling. *Global Change Biology*, **10**, 1922–1935.

De Graaff MA, Van Groenigen KJ, Six J, Hungate BA, Van Kessel C (2006) Interactions between plant growth and soil nutrient cycling under elevated CO_2: a meta-analysis. *Global Change Biology*, **12**, 2077–2091.

Dijkstra FA, Hobbie SE, Reich PB, Knops JMH (2005) Divergent effects of elevated CO_2, N fertilization, and plant diversity on soil C and N dynamics in a grassland field experiment. *Plant and Soil*, **272**, 41–52.

Downing J, Osenberg CW, Sarnelle O (1999) Meta-analysis of marine nutrient-enrichment experiments: variation in the magnitude of nutrient limitation. *Ecology*, **80**, 1157–1167.

Drake BG, Jacob J, Muehe M, Peresta G, Gonzalez-Meler M, Matamala R (1997) Acclimation of photosynthesis, respiration and ecosystem C flux of a Chesapeake Bay wetland after eight years exposure to elevated CO_2. *Plant and Soil*, **187**, 111–118.

Englund G, Sarnell O, Cooper SD (1999) The importance of data-selection criteria: meta-analyses of stream predation experiments. *Ecology*, **80**, 1132–1141.

Fung IY, Doney SC, Lindsay K, John J (2005) Evolution of C sinks in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 11201–11206.

Gill RA, Polley HW, Johnson HB, Anderson LJ, Maherali H, Jackson RB (2002) Nonlinear grassland responses to past and future atmospheric CO_2. *Nature*, **417**, 279–282.

Gurevitch J, Hedges LV (1999) Statistical issues in ecological meta-analyses. *Ecology*, **80**, 1142–1149.

Gurevitch L, Morrow LL, Wallace A, Walsh JS (1992) A meta-analysis of competition in field experiments. *American Naturalist*, **140**, 539–572.

Hedges LV, Gurevitch J, Curtis PS (1999) The meta-analysis of response ratios in experimental ecology. *Ecology*, **80**, 1150–1156.

Hedges LV, Olkin I (1985) *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL.

Hoosbeek MR, Lukac M, van Dam D *et al*. (2004) More new C in the mineral soil of a poplar plantation under Free Air C Enrichment (POPFACE): cause of increased priming effect? *Global Biogeochemical Cycles*, **18**, GB1040.

Houghton RA (2007) Balancing the global C budget. *Annual Review of Earth and Planetary Sciences*, **35**, 313–347.

Hungate BA, Jackson RB, Field CB, Chapin FS III (1996) Detecting changes in soil carbon in CO_2 enrichment experiments. *Plant and Soil*, **187**, 135–145.

Jastrow JD, Miller RM, Matamala R, Norby JR, Boutton TW, Rice CW, Owensby CE (2005) Elevated atmospheric C dioxide increases soil C. *Global Change Biology*, **11**, 2057–2064.

Jobbágy EG, Jackson RB (2000) The vertical distribution of soil organic C and its relation to climate and vegetation. *Ecological Applications*, **10**, 423–436.

Johnson DW, Ball JT, Walker RF (1997) Effects of CO_2 and nitrogen fertilization on vegetation and soil nutrient content in juvenile ponderosa pine. *Plant and Soil*, **190**, 29–40.

Johnson DW, Cheng W, Ball JT (2000) Effects of CO_2 and nitrogen fertilization on soils planted with ponderosa pine. *Plant and Soil*, **224**, 99–113.

Johnson DW, Cheng W, Joslin JD, Norby JR, Edwards NT, Todd DE (2004) Effects of elevated CO_2 on nutrient cycling in a sweetgum plantation. *Biogeochemistry*, **69**, 379–403.

Johnson DW, Curtis PS (2001) Effects of forest management on soil C and N storage: meta analysis. *Forest Ecology and Management*, **140**, 227–238.

Knorr M, Frey SD, Curtis PS (2005) Nitrogen additions and litter decomposition: a meta-analysis. *Ecology*, **86**, 3252–3257.

Lichter J, Barron SH, Bevacqua CE, Finzli AC, Irving KE, Stemmler EA, Schlesinger WH (2005) Soil C sequestration and turnover in a pine forest after six years of atmospheric CO_2 enrichment. *Ecology*, **86**, 1835–1847.

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, Merchant J (2000) *Global land cover characteristics database* (*GLCCD*), Version 2.0. Available at: http://edcdaac.usgs.gov/glcc/globdoc2_0.html.

Luo Y, Hui D, Zhang D (2006) Elevated CO_2 stimulates net accumulations of C and nitrogen in land ecosystems: a meta-analysis. *Ecology*, **87**, 53–63.

Luo Y, Reynolds JF (1999) Validity of extrapolating field CO_2 experiments to predict C sequestration in natural ecosystems. *Ecology*, **80**, 1568–1583.

Mack MC, Schuur EAG, Bret-Harte MS, Shaver GR, Chapin FS III (2004) C storage in arctic tundra reduced by long-term nutrient fertilization. *Nature*, **431**, 440–443.

Morris WF, Hufbauer RA, Agrawal AA *et al*. (2007) Direct and interactive effects of enemies and mutualists on plant performance: a meta-analysis. *Ecology*, **88**, 1021–1029.

Neff JC, Townsend AR, Gleixner G, Lehman SJ, Turnbull J, Bowman WD (2002) Variable effects of nitrogen additions on the stability and turnover of soil C. *Nature*, **419**, 915–917.

Osenberg CW, Sarnelle O, Cooper SD (1997) Effect size in ecological experiments: the application of biological models in meta-analysis. *American Naturalist*, **150**, 798–812.

Osenberg CW, Sarnelle O, Cooper SD, Holt RD (1999) Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology*, **80**, 1105–1117.

Osenberg CW, St.Mary CM, Schmitt RJ, Holbrook SJ, Chesson P, Byrne B (2002) Rethinking ecological inference: density dependence in reef fishes. *Ecology Letters*, **5**, 715–721.

Parr CS, Cummings MP (2005) Data sharing in ecology and evolution. *Trends in Ecology and Evolution*, **20**, 362–363.

Prior SA, Runion GB, Torbert HA, Rogers HH (2004) Elevated atmospheric CO_2 in agroecosystems: soil physical properties. *Soil Science*, **169**, 434–439.

Reich P, Hungate BA, Luo Y 2006 Carbon-nitrogen interactions in response to rising atmospheric CO_2. *Annual Reviews of Ecology, Evolution, and Systematics*, **37**, 611–636.

Schimel DS, House JI, Hibbard KA *et al*. (2001) Recent patterns and mechanisms of C exchange by terrestrial ecosystems. *Nature*, **414**, 169–172.

Schlesinger WH, Lichter J (2001) Limited C storage in soil and litter of experimental forest plots under increased atmospheric CO_2. *Nature*, **411**, 466–469.

Six J, Carpentier A, van Kessel C, Merckx R, Harris D, Horwath WR, Luscher A (2001) Impact of elevated CO_2 on soil organic matter dynamics as related to changes in aggregate turnover and residue quality. *Plant and Soil*, **234**, 27–36.

Smith P (2004) Monitoring and verification of soil C changes under Article 3.4 of the Kyoto Protocol. *Soil Use and Management*, **20**, 264–270.

Torbert HA, Prior SA, Rogers HH, Wood CW (2000) Review of elevated atmospheric CO_2 effects on agro-ecosystems: residue decomposition processes and soil C storage. *Plant and Soil*, **224**, 59–73.

Treseder KK, Egerton-Warburton LM, Allen MF, Cheng Y, Oechel WC (2003) Alteration of soil C pools and communities of mycorrhizal fungi in chaparral exposed to elevated C dioxide. *Ecosystems*, **6**, 786–796.

Van Groeningen KJ, Harris D, Horwath WR, Hartwig UA, van Kessel C (2002) Linking sequestration of C-13 and N-15 in aggregates in a pasture soil following 8 years of elevated atmospheric CO_2. *Global Change Biology*, **8**, 1094–1108.

Van Groenigen KJ, Six J, Hungate BA, van Kessel C, de Graaff MA, van Breemen N (2006) Element interactions limit soil carbon storage. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6571–6574.

Van Kessel C, Boots B, de Graaff MA, Harris D, Blum H, Six J (2006) Total soil C and N sequestration in a grassland following 10 years of free air CO_2 enrichment. *Global Change Biology*, **12**, 2187–2199.

Van Kessel C, Nitschelm J, Horwath WR, Harris D, Walley F, Luscher A, Hartwig U (2000) Carbon-13 input and turn-over in a pasture soil exposed to long-term elevated atmospheric CO_2. *Global Change Biology*, **6**, 123–135.

Williams MA, Rice CW, Owensby CE (2000) C dynamics and microbial activity in tallgrass prairie exposed to elevated CO_2. *Plant and Soil*, **227**, 127–137.

Xie ZB, Cadisch G, Edwards G, Baggs EM, Blum H (2005) C dynamics in a temperate grassland soil after 9 years exposure to elevated CO_2 (Swiss FACE). *Soil Biology and Biochemistry*, **37**, 1387–1395.