Sensitivity analysis and quantification of uncertainty for isotopic mixing relationships in carbon cycle research

J.M. Zobitz [a,*], J.P. Keener [a], H. Schnyder [b], D.R. Bowling [c]

[a] Department of Mathematics, University of Utah, 155 S 1400 E, Salt Lake City, UT 84112, United States
[b] Technische Universität München, Lehrstuhl für Grünlandlehre, D-85350 Freising-Weihenstephan, Germany
[c] Department of Biology, University of Utah, 257 S 1400 E, Salt Lake City, UT 84112, United States

Abstract

Quantifying and understanding the uncertainty in isotopic mixing relationships is critical to isotopic applications in carbon cycle studies at all spatial and temporal scales. Studies that depend on stable isotope approaches must also address quantification of uncertainty for parameters derived from isotopic studies. An important application of isotopic mixing relationships is determination of the isotopic content of ecosystem respiration ($\delta^{13}C_S$) via an inverse relationship (a Keeling plot) between atmospheric CO_2 concentrations ($[CO_2]$) and carbon isotope ratios of CO_2 ($\delta^{13}C$). Alternatively, a linear relationship between $[CO_2]$ and the product of $[CO_2]$ and $\delta^{13}C$ (a Miller/Tans plot) can also be applied.

We used three datasets of $[CO_2]$ and $\delta^{13}C$ in air to examine contrasting approaches to determine $\delta^{13}C_S$ and its uncertainty. These datasets were from the Niwot Ridge, Colorado, AmeriFlux site, the Biosphere-Atmosphere Stable Isotope Network (BASIN), and from the Grünschwaige Grassland Research Station in Germany. The analysis of this data included Keeling plots and Miller/Tans plots fit with both Model I (ordinary least squares) and Model II regressions (geometric mean regression and orthogonal distance regression).

Our analysis confirms previous observations that increasing the range of the measurements ($[CO_2]$ range) used for a mixing line reduces the uncertainty associated with $\delta^{13}C_S$. Using a Model II regression technique to determine $\delta^{13}C_S$ introduces a negatively skewed bias in $\delta^{13}C_S$ which is especially significant for small $[CO_2]$ ranges. This bias arises from comparatively greater variability in the dependent variable than the independent variable for a linear regression. For carbon isotope studies, uncertainty in the isotopic measurements has a greater effect on the uncertainty of $\delta^{13}C_S$ than the uncertainty in $[CO_2]$. As a result, studies that estimate parameters via a Model II regression technique maybe biased in their conclusions. In contrast to earlier studies, we advocate Model I (ordinary least squares) regression to calculate $\delta^{13}C_S$ and its uncertainty. Reducing the uncertainty of isotopic measurements reduces the uncertainty of $\delta^{13}C_S$, even when the $[CO_2]$ range of samples is small (<20 ppm). As a result, improvement in isotope (rather than $[CO_2]$) measuring capability is presently needed to substantially reduce uncertainty in $\delta^{13}C_S$. We find for carbon isotope studies no inherent advantage or disadvantage to using either a Keeling or Miller/Tans approach to determine $\delta^{13}C_S$. We anticipate that the mathematical methods developed in this paper can be applied to other applications where linear regression is utilized.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Isotopic mixing lines; Keeling plots; Ecosystem respiration; Error; Precision; Regression

1. Introduction

Understanding the balance between photosynthesis and respiration for terrestrial ecosystems represents one of the current challenges in carbon cycle research (Schimel et al., 1994, 2001; Piovesan and Adams, 2000; Valentini et al., 2000; Van Dijk and Dolman, 2004; Janssens et al., 2001). On the global scale terrestrial ecosystems uptake a large amount of carbon annually (1–2 Gt C/yr) (Prentice et al., 2001). With the increasing amount of atmospheric CO_2 from anthropogenic

* Corresponding author. Fax: +1 801 581 4148.
E-mail address: zobitz@math.Utah.edu (J.M. Zobitz).

sources coupled with current predictions of increase in global temperature, the role that forests and other biomes will play in a warmer climate is uncertain (Goulden et al., 1996; Huxman et al., 2003). On regional spatial scales, the balance of gross primary production and total ecosystem respiration can be measured as net ecosystem exchange (NEE). NEE is measured worldwide at more than 260 sites in a variety of biomes through the FLUXNET network (Baldocchi et al., 2001, http://www.fluxnet.ornl.gov). Understanding photosynthesis and respiration can help elucidate factors controlling NEE.

Isotopic mixing relationships provide a way to determine the isotopic content of net sinks or sources of ecosystem fluxes. During the night, isotopic mixing lines can determine large-scale ecosystem respiration ($\delta^{13}C_S$). Even though $\delta^{13}C_S$ is subject to biological variability (Flanagan et al., 1996; Buchmann et al., 1997a,b; Bowling et al., 2002, 2003a,b; Miller et al., 2003; Pataki et al., 2003a,b; Still et al., 2003; Lai et al., 2004; Scartazza et al., 2004; Knohl et al., 2005; Lai et al., 2005; Hemming et al., 2005), for this study we focus on the mathematical propagation of uncertainty of calculating $\delta^{13}C_S$. By comparing observations of CO_2 and $\delta^{13}C$ to simulations, it is possible to understand how errors in the ability to measure CO_2 and $\delta^{13}C$ translate into errors in $\delta^{13}C_S$. Only by understanding and working to minimize measurement uncertainty in CO_2 and $\delta^{13}C$ will it subsequently be possible to understand the effect of biological influences on $\delta^{13}C_S$. While we focus entirely on terrestrial carbon isotopic studies, we expect the results to generalize to other applications where such isotopic mixing lines are appropriate (e.g. see dietary ecology studies, Best and Schell, 1996; Burton and Koch, 1999).

$\delta^{13}C_S$ can be found through regressions of $\delta^{13}C$ and CO_2 mixing ratios ([CO_2]). A Keeling isotopic mixing line determines $\delta^{13}C_S$ as the intercept of a regression between $\delta^{13}C$ and the inverse of [CO_2] (Keeling, 1958). A Miller/Tans approach determines $\delta^{13}C_S$ as the slope of a linear regression between the product of $\delta^{13}C$ and [CO_2] versus [CO_2] (Miller and Tans, 2003). These mixing relationships will be further quantified in Section 3.2.

Pataki et al. (2003b) summarized the use of Keeling isotopic mixing lines in terrestrial carbon cycle research. Two recommendations from that study were that (a) a large sampling range of [CO_2] will minimize errors in $\delta^{13}C_S$ (see Fig. 1) and (b) Model II geometric mean regression (GMR) should be used to calculate $\delta^{13}C_S$ rather than Model I ordinary least squares regression (OLS). Fig. 1 compares errors in $\delta^{13}C_S$ versus a sample range of [CO_2] from three different
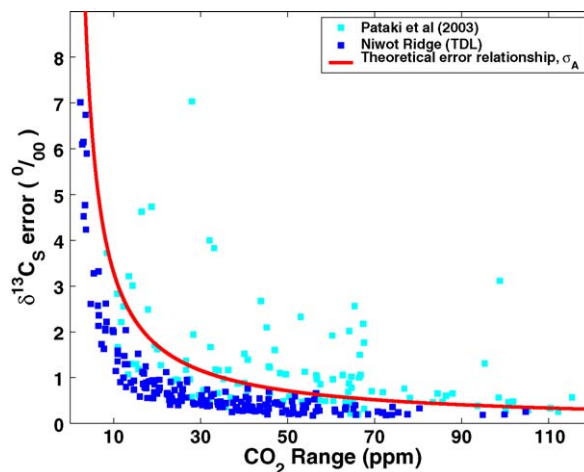

Fig. 1. Comparison of standard error of OLS intercept vs. CO_2 range from BASIN data in Pataki et al. (2003b) and this study. The BASIN data represent 146 Keeling plots from 33 sites in many biomes. The TDL study represents a single site. The solid line represents a theoretical expression for uncertainty in $\delta^{13}C_S$ vs. [CO_2] range determined from Eq. (13) developed in the text. This error line assumes that errors in [CO_2] are from a normal distribution with a standard deviation of 0.15 ppm and errors in $\delta^{13}C$ are from a normal distribution with standard deviation of 0.15‰, corresponding to the analytical uncertainty of the TDL system in Bowling et al. (2005). Notice that as the [CO_2] range decreases, the error in the intercept increases.

studies, including Pataki et al. (2003b). In general, as the sample range of [CO_2] decreases, the uncertainty on $\delta^{13}C_S$ increases.

The motivation for this study arose from inspection of large datasets of [CO_2] and $\delta^{13}C$ as collected through TDL spectrometry (Bowling et al., 2005). Such datasets have only become available recently. Because of their large size, for the first time an in-depth analysis of statistical properties of $\delta^{13}C_S$ estimates can be done. When this was done (e.g. Fig. 3), a strong bias of $\delta^{13}C_S$ estimates was revealed at low [CO_2] ranges. This bias is in contrast to previously published values of $\delta^{13}C_S$ (see panel 'b' of Fig. 3 for data from Pataki et al. (2003b)).

One highly possible explanation for the lack of a negative bias from Pataki et al. (2003b) is that the data were evaluated and screened to remove any biologically meaningless or unrealistic values of $\delta^{13}C_S$. Such quality assurance techniques include examining the standard error of the intercept or removing values with a low r^2, or coefficient of correlation. For this study we selected data near the analytical limit of the instruments and below conventional Keeling plot quality standards that would be found in the BASIN database.

Error can arise from two factors: systematic model error or analytical (measurement) error (Taylor, 1997).

Measurement error can be characterized, and, in many cases, minimized by the experimenter. Model error deals with how suitable a given model is for the data and is more difficult to quantify. For this study, we assume that the contributing factor to error in $\delta^{13}C_S$ is entirely measurement error, and not from the model assumptions underlying a mixing line.

This study examines *precision* (uncertainty or error) and *accuracy* (bias) of estimates of $\delta^{13}C_S$ obtained by Model II regression using isotopic mixing lines. To improve precision of $\delta^{13}C_S$, two protocols are to increase the number of samples taken, or in the case of field-based mass spectrometry (Schnyder et al., 2004), enhance the sample $[CO_2]$ range. While these are good experimental protocols, there are instances (e.g. Figs. 3 and 4) where a low sample range of $[CO_2]$ is obtained in spite of following the first protocol.

In the present study we examine the nature of this uncertainty and bias from a mathematical perspective in order to understand how our ability to measure $[CO_2]$ and $\delta^{13}C$ influences the calculation and interpretation of $\delta^{13}C_S$. One goal of this study was to examine if it is possible to determine $\delta^{13}C_S$ at low CO_2 ranges (<10 ppm) at an acceptable level of uncertainty in $\delta^{13}C_S$ ($<1‰$). Such determination is necessary to interpret isotopic mixing lines during the daytime. An isotopic mixing line performed during the day yields $\delta^{13}C_N$, the isotopic signature of the net exchange flux because two processes (photosynthesis and respiration) are occurring. In order to interpret daytime mixing lines, it is necessary to examine the differences between Keeling and Miller/Tans isotopic mixing lines as well as Models I and II regression techniques. As we will demonstrate, Model II regression introduces a bias in the estimation of $\delta^{13}C_S$ at small $[CO_2]$ ranges. How does this bias influence our interpretation of $\delta^{13}C_S$?

This study examines theoretical aspects of linear regression and applies the theory to characterize the calculation and statistical uncertainty of $\delta^{13}C_S$. For the theoretical aspect we (a) clarify and explain the use of Models I and II regressions and show the relationship between the two, (b) demonstrate where a bias arises in Model II regression, (c) explain the differences between a Keeling and a Miller/Tans isotopic mixing line, and (d) develop a theoretical expression for $\delta^{13}C_S$ uncertainty that describes how uncertainties in the ability to measure $[CO_2]$ and $\delta^{13}C$ influence uncertainty in $\delta^{13}C_S$ independent of regression type (Model I or II) or isotopic mixing line.

Using simulated data of $[CO_2]$ and $\delta^{13}C$, we then apply these theoretical aspects to isotopic mixing lines

by (a) comparing estimates of $\delta^{13}C_S$ to Models I and II regressions with different isotopic mixing lines (Keeling or Miller/Tans), (b) using the theoretical uncertainty expression developed to explain how the uncertainty of $\delta^{13}C$ measurements has a greater influence on the uncertainty in $\delta^{13}C_S$ than $[CO_2]$ measurements, (c) determining target levels of $\delta^{13}C$ uncertainty that will allow precise calculation of $\delta^{13}C_S$ at low $[CO_2]$ ranges, (d) recommending a regression type and isotopic mixing line for carbon cycle studies.

2. Experimental data and simulations

Direct measurements from three sources, as well as simulations, were used to examine relationships between $[CO_2]$ and $\delta^{13}C$ of CO_2:

1. *Nighttime samples*: The first, described in detail by Pataki et al. (2003b), comes from the Biosphere-Atmosphere Stable Isotope Network (BASIN) program (http://basinisotopes.org). This database includes nighttime air samples collected in 33 C_3 biomes by many different investigators and laboratories over many different years. Although different sampling methods were used, in all cases air samples were collected in flasks and returned to a laboratory for subsequent isotopic analysis via mass spectrometry. $[CO_2]$ measurements were made by infrared gas analysis or mass spectrometry in either the field or the lab.
2. *Nighttime and daytime samples*: The second dataset was collected over three consecutive summer months in a subalpine coniferous forest in Colorado, USA (the Niwot Ridge AmeriFlux site). $[CO_2]$ and $\delta^{13}C$ measurements were made by tunable diode laser absorption spectrometry (TDL) as described by Bowling et al. (2005). Measurements were made within the vegetation canopy (5, 7, 9, and 11 m height, with the canopy height 11–12 m) or near the ground (0.1, 0.5, 1, and 2 m height). Mixing lines are calculated in this paper for the two height categories separately, covering the time periods 2100–0300 h (nighttime) or 0900–1500 h (daytime). All measurements were made at the Niwot Ridge field site using the TDL. No outliers were removed from the original dataset.
3. *Daytime samples*: The third dataset was collected $[CO_2]$ and $\delta^{13}C$ data collected using a field-based mass spectrometer and instrumentation as described in Schnyder et al. (2004). The advantages to this approach include a high measurement frequency (approximately 520 $[CO_2]$ and $\delta^{13}C$ measurements

per day), and brief sampling times (0.025 s per sample) of small sample volumes (250 μl). These features increase the probability of sampling the maximum $[CO_2]$ and $\delta^{13}C$ range in a given time-window at a given site. The system is automated and obviates flask sampling and handling, and hence allows extended near-continuous measurement campaigns with minimal attendance. This dataset was obtained during a 1-week measurement campaign (3–9 September 2004) on a grazed grassland plot at Grünschwaige Grassland Research Station (Schnyder et al., 2004). In this period the system collected $[CO_2]$ and $\delta^{13}C$ measurements a few centimeters above the 7 cm tall grassland canopy approximately every 3 min. The nighttime CO_2 range was generally very large, but daytime fluctuations of CO_2 mixing ratios were small (typically less than 50 ppm in 1 h intervals). Here, we present daytime mixing lines obtained from groups of 4 or 16 pairs of $[CO_2]$ and $\delta^{13}C$ data (corresponding to measurement periods of 11 and 45 min) collected between 0900 and 1700 h local time, the brightest part of the day. No outliers were removed from the original dataset.

In addition to the observational data used, mathematical simulations were performed to investigate some aspects of mixing line methods. For these simulations, noise was added to a straight Keeling mixing relationship ($\delta^{13}C = A(1/[CO_2]) + \delta^{13}C_S$, where $A = 6567.7$ and $\delta^{13}C_S = -25.6749$). This Keeling-type relationship is further explained in Section 3.2. This relationship was arbitrarily derived from the TDL dataset to obtain realistic $[CO_2]$ and $\delta^{13}C$ values. Noise was generated by randomly sampling a normal distribution with zero mean and specified variance (described below) and adding that random sample to a particular $[CO_2]$ or $\delta^{13}C$ value on the line. Relationships (with noise added) were generated with a sample population of 1100 values, and paired subsamples of $[CO_2]$ and $\delta^{13}C$ were randomly selected from this population to create a single mixing line. We then applied Models I and II regression methods (described in Section 3.1) and the Keeling or Miller/Tans approaches to the subsamples, and repeated the entire process 10,000 times. The variances of the noise distributions added in different simulations were 0.005, 0.01, 0.1, and 0.15 ppm for $[CO_2]$ and 0.0005, 0.0075, 0.01, 0.05, 0.15‰ for $\delta^{13}C$. Typical analytical uncertainties for laboratory infrared absorption and mass spectrometer measurements on flask samples are 0.1 ppm and 0.01‰ for $[CO_2]$ and $\delta^{13}C$, respectively (Miller and Tans, 2003). Uncertainties for TDL

measurements are typically 0.15 ppm and 0.15‰ (Bowling et al., 2005).

3. Theory: determining $\delta^{13}C_S$ from measurements

In this section we describe linear regression methods, isotopic mixing lines used, and discuss uncertainty in $\delta^{13}C_S$.

3.1. Linear regression formulas

When fitting experimental observations to a theoretical relationship, one needs to consider the purpose of the regression. *Predictive regression* utilizes a relationship from measured data to determine the dependent variable for a measured independent variable. *Functional regression* estimates parameters from an assumed functional relationship between measured values. One goal of functional regression is to compare the same parameters estimated from two different datasets (Ricker, 1973; Sprent and Dolby, 1980).

A linear regression seeks to find the best-fit line ($y = a + bx$) that can be fit through a collection of data points (x_i, y_i). As shown in Fig. 2 there are at least three conceptual ways to determine this best-fit line. By practicality we would expect that different regression techniques yield consistent parameter estimates. The most common type of linear regression is Model I regression, known as ordinary least squares (OLS). OLS assumes a definite relationship between the independent (x) and dependent (y) variables with no variation in the independent variable. OLS seeks to minimize the sum of the squared residuals from the hypothetical best-fit line to each data point. Using standard calculus techniques, one can find that the a and the b that minimize the residual are given by

$$b_{OLS} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \qquad a_{OLS} = \bar{y} - b_{OLS}\bar{x},$$

(1)

where an overbar represents the mean of the data (Taylor, 1997; Laws, 1997). Model I regression is appropriate for predictive purposes (Ricker, 1973).

If there is no preferred independent variable between data x and y, Model II regression is used to reduce the influence of forcing a deterministic relationship between data. Instead of minimizing just the vertical residuals through data points, Model II regression minimizes the vertical and horizontal residuals as well. There are many different ways one can minimize the

Ordinary Least Squares (OLS)

(a)

Geometric Mean Regression (GMR)
(Two regressions performed)

(b)
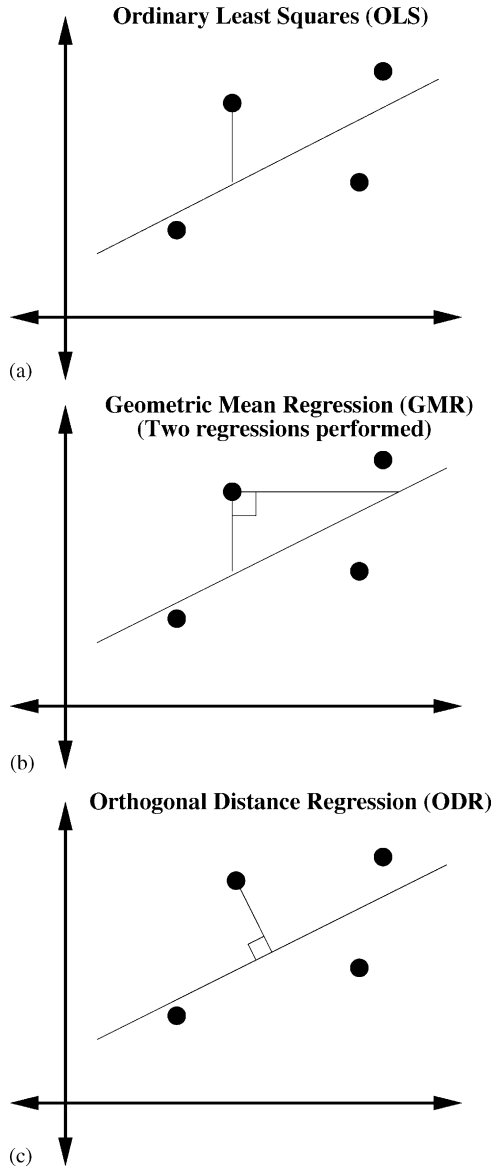
Orthogonal Distance Regression (ODR)

(c)

Fig. 2. Conceptual comparison of regression methods and minimization of residual, or the distance from a data point to the best-fit line. A Model I regression is ordinary least squares (OLS), which minimizes the vertical residual. Model II regression techniques include geometric mean regression (GMR), which incorporates the minimum of the vertical and horizontal residual from the best-fit line, and orthogonal distance regression (ODR), which minimizes the perpendicular distance of the residual from the best-fit line. Note that the final values of the best-fit line coefficients for each method may be different.

sum of the vertical and horizontal residuals. Two types commonly used in the literature are geometric mean regression (GMR) (Ricker, 1973; Laws, 1997) or orthogonal distance regression (ODR) (Laws, 1997).

GMR was recommended by Flanagan et al. (1996) as well as Pataki et al. (2003b) for determining $\delta^{13}C_S$ using Keeling plots because there is no clear specification of an independent variable (e.g. $[CO_2]$ does not produce a $\delta^{13}C$ measurement and vice versa). Geometric mean regression can be derived by rearranging the equation $y = a + bx$ to $x = (1/b)y - (a/b)$ and performing an OLS regression with y as the independent variable. As a result, note that doing the regression with y as the independent variable, the slope is equal to the reciprocal of an OLS regression with x as the independent variable. Thus, we have two candidates for the slope, and we take the geometric mean of them to find b and subsequently a:

$$b_{GMR} = \sqrt{b_{yx}\frac{1}{b_{xy}}}, \qquad a_{GMR} = \bar{y} - b_{GMR}\bar{x}, \qquad (2)$$

where b_{yx} is the slope obtained by doing a regression with x as the independent variable and b_{xy} is the slope obtained by doing a regression with y as the independent variable. The coefficient b_{GMR} can be rewritten in terms of the correlation coefficient r, where r is equal to (Taylor, 1997):

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}. \qquad (3)$$

Doing this we see that

$$
\begin{aligned}
b_{GMR} &= \sqrt{b_{yx}\frac{1}{b_{xy}}} = \sqrt{b_{yx}\frac{b_{yx}}{b_{yx}}\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}} \\
&= b_{yx}\sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}} \\
&= b_{yx}\frac{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})} \\
&= \frac{b_{yx}}{r} = \frac{b_{OLS}}{|r|}. \qquad (4)
\end{aligned}
$$

The absolute value on r is necessary in order for b_{GMR} to have the same sign as b_{OLS}. Eq. (4) will be come important later when examining the relative merits of each regression method (e.g. Section 4).

Using planar trigonometry the appropriate residual for orthogonal distance regression can be calculated, and when done so, the coefficients for

the best-fit line obtained are (Boggs and Rogers, 2001; Laws, 1997):

$$b_{ODR} = -B \pm \sqrt{B^2 + 1},$$

$$\text{where } B = \frac{1}{2} \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2 - (y_i - \bar{y})^2}{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}, \tag{5}$$

$$a_{ODR} = \bar{y} - b_{ODR}\bar{x}. \tag{6}$$

3.2. Isotopic mixing lines

To determine $\delta^{13}C_S$, the isotopic signature of ecosystem respiration, a widely used approach is the Keeling plot method (Keeling, 1958). Assume that there are only two pools of carbon dioxide, a net source or sink pool and a background pool. For notational convenience let C_X be equal to CO_2 from pool X and let $\delta^{13}C_X$, the isotopic signature of that pool, be denoted by δ_X.

By mass conservation:

$$C_{Obs} = C_B + C_S, \tag{7}$$

where C_{Obs} is the total measured CO_2, C_S the pool of CO_2 influenced by biological sources (photosynthesis and respiration), and C_B is the background CO_2. Similarly, conservation of $^{13}CO_2$ is given by the following equation:

$$\delta_{Obs}C_{Obs} = \delta_B C_B + \delta_S C_S, \tag{8}$$

See Keeling (1958) for the derivation of Eq. (8). Substituting Eq. (7) into Eq. (8) and rearranging terms, one obtains the usual Keeling relationship:

$$\delta_{Obs} = \frac{C_B(\delta_B - \delta_S)}{C_{Obs}} + \delta_S. \tag{9}$$

Multiplying through by C_{Obs} in Eq. (9), one obtains an alternative regression which we refer to as the Miller/Tans regression (Miller and Tans, 2003):

$$\delta_{Obs}C_{Obs} = C_B(\delta_B - \delta_S) + \delta_S C_{Obs}. \tag{10}$$

Eqs. (9) and (10) are nonlinear in $[CO_2]$ and $\delta^{13}C$, but under the appropriate transformations $(x = 1/C_{Obs}, y = \delta_{Obs}$ for Keeling, $x = C_{Obs}, y = C_{Obs}\delta_{Obs}$ for Miller/Tans), they become linear. For a Keeling regression, $\delta^{13}C_S$ is the y-intercept of a regression between the transformed data, and for a Miller/Tans regression, $\delta^{13}C_S$ is the slope.

It is known that $[CO_2]$ is negatively correlated with values of $\delta^{13}C$, so the inverse of $[CO_2]$ in a Keeling regression is *positively* correlated with $\delta^{13}C$. With a similar argument we expect that the slope of a Miller/Tans regression should be negative.

3.3. Theoretical expression of uncertainty in $\delta^{13}C_S$

We wish to examine the uncertainty in $\delta^{13}C_S$ associated with various methods of calculating it. Pataki et al. (2003b) recommend using GMR (a Model II regression technique) to determine $\delta^{13}C_S$ but also recommend using the standard error of the intercept and slope of a Model I regression to determine uncertainty. A formula for the standard error of the OLS intercept and slope can be found in standard statistical texts (Taylor, 1997). A criticism of using a Model I uncertainty estimate for Model II regression is that since Model II regression accounts for errors in both the independent and dependent variables, using a Model I technique may not completely characterize the uncertainty. What if the standard error of the intercept *overestimates* the uncertainty on $\delta^{13}C_S$? Furthermore, the standard error of the intercept provides little insight into how measurement or sampling errors might influence uncertainty in $\delta^{13}C_S$ (Miller and Tans, 2003). As stated in Section 1, we assume that systematic error is negligible.

To investigate the influence of measurement errors of $[CO_2]$ and $\delta^{13}C$ on uncertainty of $\delta^{13}C_S$, we develop here a theoretical relationship between measurement error of $[CO_2]$ and $\delta^{13}C$ and investigate how that translates to error in $\delta^{13}C_S$. The importance of such an expression is that it allows one to directly investigate the effect of uncertainty in the measured variables ($[CO_2]$ and $\delta^{13}C$) on the uncertainty of $\delta^{13}C_S$. We represent measurement errors on $[CO_2]$ or $\delta^{13}C$ as random variables. In effect what is measured is a characterization of the "true" value with some error. It is also possible to characterize the various values the error may take on as a *distribution*. For a random variable Z with a distribution, two key parameters that characterize it are its expected value (or mean) and its standard deviation. Denote the expected value of Z by $\langle Z \rangle$ and its standard deviation by σ_Z.

Assume our measurements of $[CO_2]$ and $\delta^{13}C$ are nonlinearly related, either through a Keeling or a Miller/Tans isotopic mixing line. Since $\delta^{13}C_S$ is a parameter estimated from a mixing line, it will also be a random variable with a distribution. In particular, we wish to determine the expected value and variance of the distribution of $\delta^{13}C_S$.

To conduct an uncertainty analysis on the dataset, observe that the slope (b) or intercept (a) in Eqs. (1), (2) and (5) *all* depend continuously on the data (x_i, y_i). Assume that both the independent and dependent variables have errors that are normally distributed with mean 0 and standard deviation σ. As a result, a particular measurement $x_i = X_i - \epsilon_i$, where ϵ_i is a sample of a normal distribution with variance σ_x^2, and X_i represents the true measured value without error. A similar statement holds for y_i: $y_i = Y_i - \eta_i$. Denote the set of n measurements of x by $\vec{x}$, the set of n measurements of y by $\vec{y}$, the term $\vec{\epsilon}$ is the vector of errors for the variable $\vec{X}$, and $\vec{\eta}$ is the vector of errors for the variable $\vec{Y}$. $\vec{\epsilon}$ and $\vec{\eta}$ are both Gaussian random variables with zero mean and standard deviation σ_X and σ_Y, respectively. Note that the expected value of $\vec{\epsilon}$ and $\vec{\eta}$ are both zero by definition.

Assuming that the particular errors involved are small, we can expand a or b from our linear regression formula in a Taylor series about the point (X_i, Y_i) and neglect the higher order terms (denoted by $O(\epsilon^2 + \sigma^2)$). This is an n dimensional Taylor series expansion. For the following, A will denote either the intercept (a) or slope (b):

$$A(\vec{x}, \vec{y}) = A(\vec{X} - \vec{\epsilon}, \vec{Y} - \vec{\eta}) \approx A(\vec{X}, \vec{Y})$$
$$+ \sum_{i=1}^{n} \frac{\partial A(X_i, Y_i)}{\partial X_i} \epsilon_i + \sum_{i=1}^{n} \frac{\partial A(X_i, Y_i)}{\partial Y_i} \eta_i$$
$$+ O(\epsilon^2 + \eta^2) \approx A(\vec{X}, \vec{Y})$$
$$+ \nabla_x A(X_i, Y_i) \cdot \vec{\epsilon} + \nabla_y A(X_i, Y_i) \cdot \vec{\eta}, \quad (11)$$

where

$$\sum_{i=1}^{n} \frac{\partial A(X_i, Y_i)}{\partial X_i} \epsilon_i = \nabla_x A(X_i, Y_i) \cdot \vec{\epsilon},$$
$$\sum_{i=1}^{n} \frac{\partial A(X_i, Y_i)}{\partial Y_i} \eta_i = \nabla_y A(X_i, Y_i) \cdot \vec{\eta},$$

The expression $A(\vec{x}, \vec{y})$ is a random variable. Denote its expected value as $\langle A(\vec{x}, \vec{y}) \rangle$, which can be found using standard statistical techniques (Gubbins, 2004):

$$\langle A(x_i, y_i) \rangle = \langle A(X_i, Y_i) + \nabla_x A(X_i, Y_i) \cdot \vec{\epsilon}$$
$$+ \nabla_y A(X_i, Y_i) \cdot \vec{\eta} \rangle = A(X_i, Y_i)$$
$$+ \langle \nabla_x A(X_i, Y_i) \cdot \vec{\epsilon} \rangle + \langle \nabla_y A(X_i, Y_i) \cdot \vec{\eta} \rangle$$
$$= A(X_i, Y_i) + \nabla_x A(X_i, Y_i) \cdot \langle \vec{\epsilon} \rangle$$
$$+ \nabla_y A(X_i, Y_i) \cdot \langle \vec{\eta} \rangle = A(X_i, Y_i). \quad (12)$$

Eq. (12) states that even though $\vec{x}$ and $\vec{y}$ have measurement error, the expected value of the distribution is the value that A would be if $\vec{x}$ and $\vec{y}$ have no measurement error. The variance σ_A^2 of $A(x_i, y_i)$ is defined as the expected value of

$$[A(x_i, y_i) - A(X_i, Y_i)]^2.$$

Assume that errors $\vec{\epsilon}$ and $\vec{\eta}$ are independent and uncorrelated. Carrying out this calculation in much the same manner as above, it can be shown that

$$\sigma_A^2 = \left[\sum_{i=1}^{n} \left(\frac{\partial A(X_i, Y_i)}{\partial X_i} \right)^2 \right] \sigma_X^2$$
$$+ \left[\sum_{i=1}^{n} \left(\frac{\partial A(X_i, Y_i)}{\partial X_i} \right)^2 \right] \sigma_Y^2 \quad (13)$$

$$= \alpha(\vec{X}, \vec{Y}) \sigma_X^2 + \beta(\vec{X}, \vec{Y}) \sigma_Y^2, \quad (14)$$

If there were correlation between errors $\vec{\epsilon}$ and $\vec{\eta}$ (i.e. $\langle \epsilon_i \epsilon_j \rangle \neq 0$ or $\langle \eta_i \eta_j \rangle \neq 0$ for $i \neq j$, or $\langle \eta_i \epsilon_j \rangle \neq 0$ for all i and j), then additional terms in Eq. (13) would arise, but these covariance terms would not affect Eq. (12).

The main result of this analysis is that the variance in either of the regression parameter a or b is a linear combination of the variances on x and y. The terms $\alpha(\vec{X}, \vec{Y})$, $\beta(\vec{X}, \vec{Y})$ (or for simplicity α and β) are the *error reduction/amplification factors*. These are dependent on the following:

- The values of the true data (X_i, Y_i). These values certainly will not be known a priori, however if one knows the prior measurement errors σ_x and σ_y, we recommend subtracting from the data errors that have been drawn from a normal distribution a mean 0 and variance σ^2 to obtain a best approximation to α and β.
- The type of regression used (OLS, GMR, ODR). For each regression, the coefficients a and b have explicit formulas. The error amplification factors α and β in Eq. (13) can be quite complicated to determine.
- The isotopic mixing line used (Keeling or Miller/Tans). Calculating the error reduction factors involves differentiating a vector. In Appendix A, we show how to calculate the sensitivity factors for a Miller/Tans OLS regression.
- For carbon isotope studies, we can now conclude that to first order in Eq. (13), measurement uncertainties in the measured variables [CO_2] and $\delta^{13}C$ are linearly related to uncertainty in $\delta^{13}C_S$, irrespective of whether a Keeling or a Miller/Tans plot is used. What could potentially distinguish a Keeling from a Miller/Tans plot are different error reduction/amplification factors.
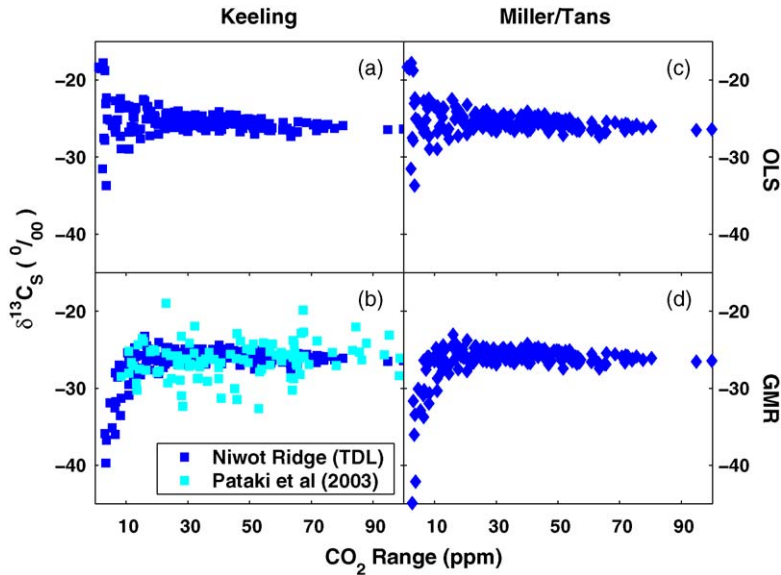- Eq. (13) provides a general theoretical estimate of uncertainty in $\delta^{13}C_S$. This is a useful point for

Fig. 3. Nighttime Keeling and Miller/Tans $\delta^{13}C_S$ for each of the regression schemes (OLS, GMR, ODR). Data from Pataki et al. (2003b) are included with Keeling GMR for comparison. Note the negative bias associated with the Model II (GMR, ODR) regression methods. $\delta^{13}C_S$ for the present study was calculated from nighttime TDL data by grouping inlets located at 5, 7, 9, and 11 m together and 0.1, 0.5, 1 and 2 m together in a 6 h interval centered on midnight (Bowling et al., 2005). Each point on the figure for the Niwot Ridge study represents a mixing line with $n \approx 20$ samples of [CO_2] and $\delta^{13}C$.

comparison to the standard error of the OLS intercept recommended by Pataki et al. (2003b).

With Eq. (13) we now have an expression for the uncertainty of $\delta^{13}C_S$. Ignoring model error, this expression is dependent on the analytical uncertainty in measured quantities and the type of regression and mixing line used to calculate $\delta^{13}C_S$. Because of its generality, we can investigate how analytical uncertainty influences $\delta^{13}C_S$ uncertainty on all the different possible combinations used to calculate $\delta^{13}C_S$ (OLS, GMR, or ODR paired with Keeling or Miller/Tans mixing lines).

4. Discussion

4.1. Evaluation of regression methods

Bowling et al. (2005) demonstrated that there was a negative bias in GMR Keeling plot determinations of $\delta^{13}C_S$ when the range of [CO_2] is small. Shown in Fig. 3 are nighttime calculations of $\delta^{13}C_S$ versus [CO_2] range for OLS and GMR regressions for both Keeling or Miller/Tans isotopic mixing lines from   data collected by the TDL. For comparison, panel 'b' of Fig. 3 features BASIN data from Pataki et al. (2003b) with variable precision from each measurement site. The BASIN data are provided as a point of comparison

to the TDL. As we explain below, BASIN data should not be assumed to have equal measurement precision to the TDL, and thus comparing the two is misleading. Panels 'b' and 'd' in Fig. 3 demonstrate that GMR negatively biases $\delta^{13}C_S$ if either a Keeling or Miller/Tans mixing line is used. Plots of ODR produced similar results as GMR and are not shown.

The bias is more pronounced when mixing lines are performed using daytime measurements. Fig. 4 is structured similarly to Fig. 3, but presents daytime results for the mixing lines. Daytime ranges for the mixing ratios collected by the TDL are smaller than nighttime ranges, leading to greater variability in $\delta^{13}C_S$. Panel 'd' of Fig. 4 compares TDL data to data from the field-based mass spectrometer at Grünschwaige. Furthermore, note that in Fig. 4 GMR negatively biases $\delta^{13}C_S$.

The three datasets represent very different methods of collection, yet Figs. 3 and 4 suggest that determining $\delta^{13}C_S$ from a Keeling or a Miller/Tans regression give consistent results. (This point is discussed in Section 4.4 and Fig. 11.) One way to circumvent the negative bias is to increase the probability of sampling maximum range of [CO_2] and $\delta^{13}C$ for a given time window, as done in the Grünschwaige study in Fig. 5. However, note that even when this is done, the possibility for bias at low [CO_2] ranges *still* exists. Furthermore, Fig. 5 demonstrates that increasing the number of samples for a
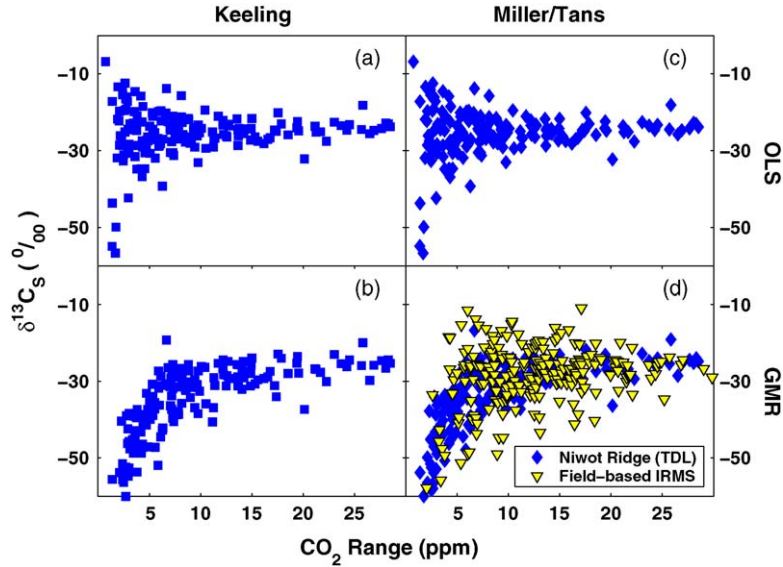
Fig. 4. Daytime Keeling and Miller/Tans $\delta^{13}C_S$ for each of the regression schemes (OLS, GMR, ODR). Note the negative bias associated with the Model II (GMR, ODR) regression methods. Data from the Grünschwaige field-based IRMS are included for comparison (Schnyder et al., 2004). Each $\delta^{13}C_S$ from the Grünschwaige field-based IRMS represents a mixing line with four pairs of [CO_2] and $\delta^{13}C$ data. $\delta^{13}C_S$ for the present study was calculated from daytime TDL data by grouping inlets located at 5, 7, 9, and 11 m together and 0.1, 0.5, 1 and 2 m together in a 6 h interval centered on 1200 h (Bowling et al., 2005). Each point on the Figure for the Niwot Ridge study represents a mixing line with $n \approx 20$ samples of [CO_2] and $\delta^{13}C$.
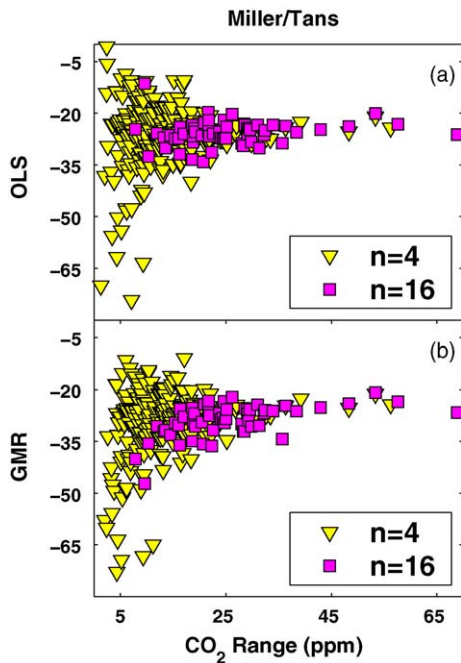

Fig. 5. Daytime Miller/Tans $\delta^{13}C_S$ calculated via OLS (panel a) or GMR (panel b) using a field-based mass spectrometer as described in Schnyder et al. (2004). Each point represents a mixing line with either 4 (triangles) or 16 pairs (squares) of [CO_2] and $\delta^{13}C$ data.

mixing line from $n = 4$ or $n = 16$ reduces the effect of the bias, but yet it is still present in panel 'b' of Fig. 5. We discuss the effect of increasing the number of samples on $\delta^{13}C_S$ estimates in Section 4.4.

The BASIN data do not show a bias as strong as the Niwot Ridge data or Grünschwaige data. We randomly
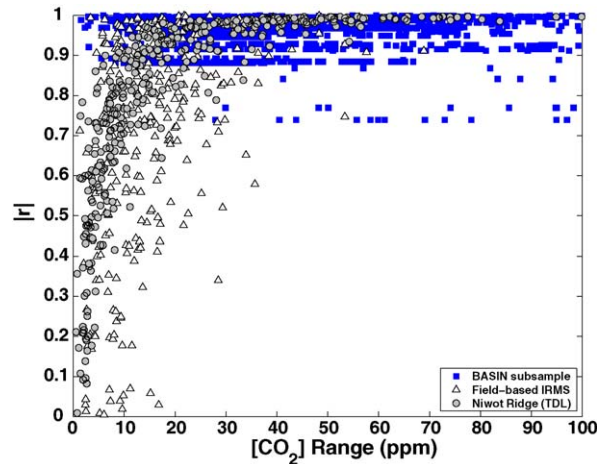

Fig. 6. Comparison of correlation coefficient r to [CO_2] range for the BASIN subsampling described in text (squares), the Grünschwaige field-based IRMS (triangles) (Schnyder et al., 2004), and the Niwot Ridge study (circles) (Bowling et al., 2005). Note the sharp drop in |r| values at low [CO_2] ranges for the Grünschwaige and Niwot Ridge studies, but not for the BASIN subsampling.
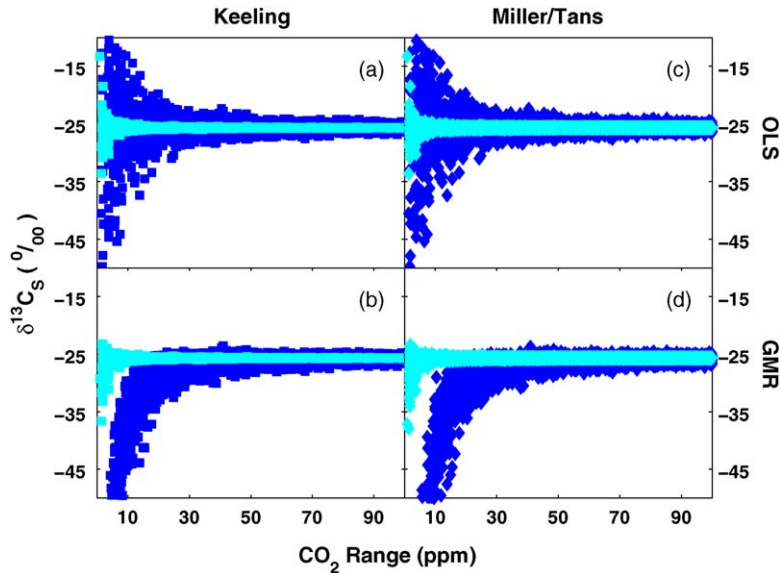
Fig. 7. Simulation results for $\delta^{13}C_S$ vs. [CO_2] range for each of the regression schemes (OLS, GMR, ODR) and mixing relationships (Keeling or Miller/Tans) using a straight mixing line that has been randomly perturbed by 0.15 ppm in CO_2 and 0.15‰ in $\delta^{13}C$ (dark symbols) or 0.15 ppm and 0.01‰ (light symbols). The correct $\delta^{13}C_S$ (with no error added) is −25.67‰. Each mixing line had 20 samples. Note the significant reduction in uncertainty for the 0.01‰ data.

subsampled each BASIN site [CO_2] and $\delta^{13}C$ dataset to see if a bias could be produced. We randomly picked at least four subsamples from the data for each site and computed $\delta^{13}C_S$ for each of the three regression methods fit with both Keeling or Miller/Tans mixing lines. After 250 subsamples of each dataset for a site, a strong bias (as shown in Figs. 3 and 4) was not produced (results not shown). However, we can attribute this lack of a bias to the fact that each dataset was likely screened before publication to remove erroneous results. Eq. (4) shows that as the coefficient of correlation r decreases, the estimate of b_{GMR} is biased relative to b_{OLS}. Fig. 6 shows a plot of r versus [CO_2] range for the BASIN subsampling, Niwot Ridge, and Grünschwaige studies. Even when the range of samples is small, the BASIN data still exhibit high r values. This may be attributable to many factors: processing of $\delta^{13}C$ samples at a higher level of precision compared to the TDL or field-based mass spectrometry, prescreening of data to remove outliers, or rejection of datasets with poor r values. Due to this higher apparent precision of the BASIN data, it is misleading to compare the BASIN data to the Niwot Ridge data as done in panel 'b' of Fig. 3. Fig. 6 underscores the strength of the BASIN data and the conclusions of Pataki et al. (2003b) in spite of the mathematical challenges underlying Model II regression. Furthermore, Fig. 6 demonstrates the usefulness of high-

resolution measurements to investigate basic assumptions about data (e.g. Model I versus Model II, Keeling versus Miller/Tans) that would not be possible with flask-based measurements or instrumentation.

To investigate whether this bias arose due to problems with the datasets, we simulated mixing lines with artificial data as described in Section 2. These results are presented in Figs. 7 and 8. The dark symbols in Fig. 7 represent a dataset of [CO_2] and $\delta^{13}C$ that has been perturbed with normally distributed random error with zero mean and standard deviation of 0.15 ppm, 0.15‰, respectively. The light symbols represent a dataset perturbed with 0.15 ppm, 0.01‰ error, respectively. The correct $\delta^{13}C_S$ (with no error added) is −25.67‰. The same trends in Figs. 3 and 4 emerge in Fig. 7: Model II negatively biases $\delta^{13}C_S$ at low [CO_2] ranges and using a Keeling or Miller/Tans regression give consistent estimates of $\delta^{13}C_S$. Estimates of $\delta^{13}C_S$ for high [CO_2] ranges converge on the "correct" value of $\delta^{13}C_S$ with no error added for any regression or mixing line. As a result, we expect large sample ranges to lead to consistent parameter estimates for either Model I or II regression. While decreasing the random error in $\delta^{13}C$ seems to reduce the variability in $\delta^{13}C_S$ at low [CO_2] ranges, Fig. 8 shows that the variability is still present, although the [CO_2] range where this variability dominates is smaller. These results indicate that estimating $\delta^{13}C_S$ at low [CO_2] ranges depends on
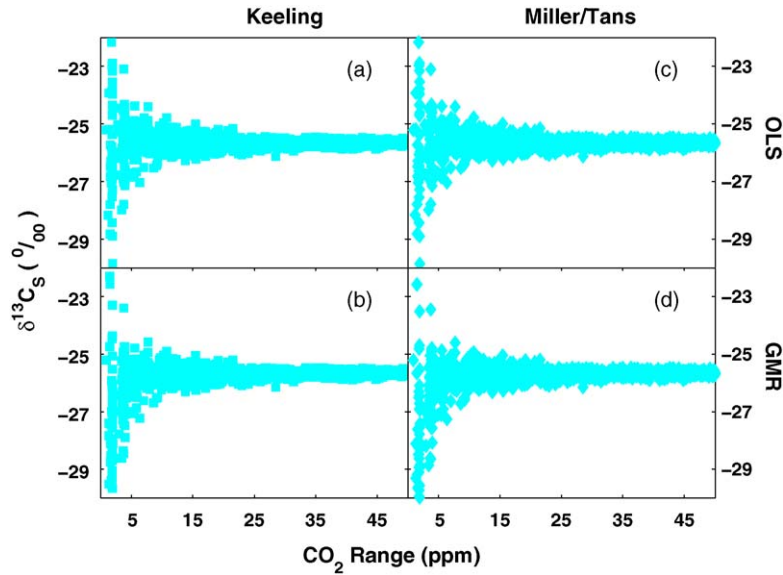
Fig. 8. Same data as gray symbols in Fig. 7 but with changed scales.

the ability to measure [CO$_2$] and δ^{13}C with a minimum amount of measurement or sampling error rather than a particular choice of isotopic mixing line (Keeling or Miller/Tans).

4.2. Explaining Model II regression bias

From Eq. (4) we see that as data become less correlated (as $|r|$ approaches 0), then $b_{OLS} \neq b_{GMR}$. In particular as the correlation coefficient goes to 0, b_{GMR} will diverge to positive or negative infinity, depending on the sign of b_{OLS}. (For δ^{13}C$_S$ it will be negative infinity.) In effect, geometric mean regression *biases* the regression coefficients *a* and *b* at low $|r|$. This bias is apparent in Figs. 3–5 for *both* Keeling and Miller/Tans regressions calculated with GMR as well as ODR (ODR results not shown). Furthermore, because the slopes for b_{GMR} and b_{ODR} can be related to b_{OLS} (Eqs. (1)–(5)), we expect that any variability in b_{OLS} at low [CO$_2$] ranges will lead to variability in b_{ODR} and b_{GMR} (similarly for a_{GMR} and a_{ODR}). In spite of this variability, are the estimates of b_{GMR} and b_{ODR} (a_{GMR} and a_{ODR}) consistent or unbiased with respect to b_{OLS} (a_{OLS})?

To further investigate the reasons for this bias, one can scale the regression axes appropriately so that the independent and dependent variables for a particular regression are scaled between -1 and 1 and have zero mean and unit variance. This is necessary for our application because for a given mixing line, there is a mismatch between the signal to noise ratios (SNR) for a [CO$_2$] and δ^{13}C measurement, respectively. This

nondimensionalization is done by the following transformation:

$$x \rightarrow x^* = \frac{x - \bar{x}}{\sigma_{x_{data}}}, \qquad y \rightarrow y^* = \frac{y - \bar{y}}{\sigma_{y_{data}}} \qquad (15)$$

The result of this transformation is that one variable will not unduly influence the regression coefficients *a* and *b* because the variables are dimensionless (Gubbins, 2004). Note that in Eq. (15), $\sigma_{x_{data}}$ is the standard deviation of all the *x* samples on a particular mixing line, which will usually be different from the analytical uncertainty in an individual measurement. Denote the intercept and slope from a fitted relationship in the (x^*, y^*) dimensionless coordinate system as a^* and b^*, respectively. When this is done, translating back to the original dimensional coordinate system (x, y), the *a* and *b* are scaled accordingly:

$$b = \frac{\sigma_{y_{data}}}{\sigma_{x_{data}}} b^* \qquad (16)$$

$$a = \sigma_{y_{data}} a^* + \bar{y} - b\bar{x} \qquad (17)$$

Under the assumption that the errors in the dependent variable are normally distributed, then the coefficients a^* and b^* for OLS will then be normally distributed (Taylor, 1997; Tarantola, 2005). This property for OLS will also be preserved when transforming back into the dimensional coordinate system. However, in the (x^*, y^*) coordinate system both axes are symmetrical, and so a

fit of y^* versus x^* will yield the same slope as a regression with x^* as the dependent variable. As a result, for geometric mean regression, b^*_{GMR} will equal 1. Similarly, x^* and y^* in this coordinate system have a variance of 1, so B in Eq. (5) will be zero, giving $b^*_{ODR} = 1$. For a Miller/Tans mixing line, $b^* = -1$ for GMR and ODR. As the (x^*, y^*) coordinate system is scaled between -1 and 1, one should expect that a^* be 0. For both Model II regressions, simulations (not shown) using artificial data (described in Section 2) confirmed that a^* was normally distributed about 0.

The bias associated with Model II regression arises due to the term $\sigma_{y_{data}}/\sigma_{x_{data}} = \theta$ in Eq. (16). Fig. 9 presents simulation results of $\sigma_{y_{data}}/\sigma_{x_{data}}$ for both Keeling and Miller–Tans regressions from an artificial dataset with 0.15 ppm, 0.15‰ random error. Notice how this ratio tends to positive infinity as $[CO_2]$ sample range decreases. When translating back into the dimensional coordinate system, this ratio biases the regression coefficient a or b.

Model II estimates of b (and concurrently a) in the dimensioned space will have an increasing bias as $[CO_2]$ range decreases. Since $b^* = \pm 1$ for Model II regression, b will diverge to negative infinity as $[CO_2]$ range decreases. Recall that for a Miller/Tans mixing line, $b = \delta^{13}C_S$. For a Keeling mixing line, a in Eq. (17) is weighted by the biased factor $b\bar{x}$ and will likewise bias a to negative infinity. Since a in a Keeling mixing line is $\delta^{13}C_S$, then $\delta^{13}C_S$ will have a negative bias.

Contrast this result to Model I estimates of a and b. The factor $\sigma_{y_{data}}/\sigma_{x_{data}}$ when translating back to the (x, y) coordinate system is still relevant for Model I regression, however examination of Eq. (1) shows that b^* will *not* be 1, and will vary as $[CO_2]$ range decreases, leading to a normal distribution of b^*.

As a result, the bias associated with Model II regression is mathematical in nature. If GMR or ODR is used to estimate $\delta^{13}C_S$, they will bias $\delta^{13}C_S$ particularly at low r (correlation coefficient). Furthermore, this bias arises independent of a particular mixing line choice (Keeling or Miller/Tans).

We acknowledge that the bias arising from the term $\sigma_{y_{data}}/\sigma_{x_{data}}$ in Eq. (16) is specific to the particular data and study at hand. When this ratio was calculated for the BASIN subsampling scheme (described in Section 4.1), $\sigma_{y_{data}}/\sigma_{x_{data}}$ was constant at low $[CO_2]$ ranges (results not shown). Furthermore, we do not exclude the possibility that Model I regression *may be* biased at low ranges in the independent variable if there is a consistent covariance between the regression variables (i.e. the term $\sum x_i^* y_i^*$ in Eq. (1) is consistently positive or negative at all ranges). However, for the mixing line studies examined here, investigation of mixing lines at low $[CO_2]$ ranges showed mixing lines that would produce slopes (and hence intercepts) that were opposite in sign to our mixing line assumptions in Section 3.2. This random covariance at low $[CO_2]$ ranges was also reflected in our simulated data (Fig. 7). We argue that the mixing line relationships developed should hold at small $[CO_2]$ ranges, but our ability to measure these relationships at low ranges breaks down because of a small signal to noise ratio (Table 2, discussed below). Thus, for current analytical error levels we should not expect a bias for OLS in mixing line studies.

4.3. Determining $\delta^{13}C_S$ at low $[CO_2]$ ranges

Fig. 7 shows results of $\delta^{13}C_S$ versus $[CO_2]$ range from two simulated datasets with normally distributed random errors with 0.15 ppm, 0.15‰ and 0.15 ppm, 0.01‰ standard deviations, respectively. For each simulation, the uncertainty in $\delta^{13}C_S$ was calculated for each particular regression type (OLS and GMR) and
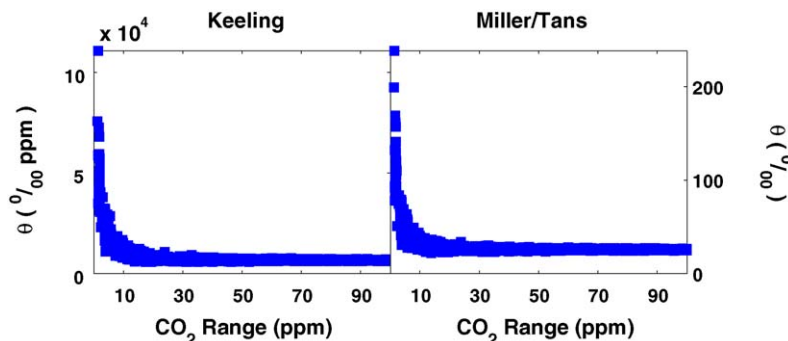
Fig. 9. Comparison of the ratios of $\theta = \sigma_{y_{data}}/\sigma_{x_{data}}$ for Keeling and Miller/Tans regressions vs. $[CO_2]$ range using artificial data that has 0.15 ppm, 0.15‰ random error. Note the different dimensional units on θ for a Keeling or a Miller/Tans regression. Also note that as $[CO_2]$ range decreases, both of these ratios diverge to infinity, explaining the reason for bias in Model II regressions.

Table 1
Mean, median, and skewness for each of the distributions of $\delta^{13}C_S$ generated from simulations shown in Fig. 7

Dataset	Regression	Mean (‰)	Median (‰)	Median − mean (‰)	Skewness	Mean (‰)	Median (‰)	Median − mean (‰)	Skewness
A	kOLS	−25.62	−25.67	−0.05	5.94	−25.67	−25.68	−0.01	0.59
A	kGMR	−27.60	−26.07	1.53	0.73	−26.48	−26.03	0.45	−2.73
A	mtOLS	−25.62	−25.67	−0.05	5.93	−25.67	−25.67	0.00	0.49
A	mtGMR	−33.18	−26.14	7.04	−31.92	−26.57	−26.07	0.50	−2.69
B	kOLS	−25.67	−25.67	0.00	5.94	−25.67	−25.67	0.00	5.62
B	kGMR	−25.71	−25.67	0.04	0.73	−25.71	−25.68	0.03	−17.91
B	mtOLS	−25.67	−25.67	0.00	5.93	−25.67	−25.67	0.00	5.63
B	mtGMR	−25.72	−25.68	0.04	−31.92	−25.72	−25.68	0.04	−26.78

Simulations from dataset "A" represent data with normally distributed random errors with 0.15 ppm, 0.15‰ standard deviations. Simulations from dataset "B" represent data with normally distributed random errors with 0.15 ppm, 0.01‰ standard deviations. The column "regression" specifies the type of regression (OLS or GMR) and mixing line used (k = Keeling, mt = Miller/Tans). For ODR, similar results to GMR were obtained and hence omitted. The left columns represent the mean, median, and skewness of the entire population of $\delta^{13}C_S$ from a particular regression. The right columns report the mean, median, and skewness for the population of $\delta^{13}C_S$ restricted to be within 10‰ of −25.67‰, the correct $\delta^{13}C_S$ with no error added.

isotopic mixing line (Keeling or Miller/Tans) from Eq. (13). For ODR, similar results to GMR were obtained and hence omitted.

If a distribution is normally centered about a mean, then the mean will be equal to the median of that distribution. If a distribution is negatively skewed, then the mean will be less than the median, implying the mean subtracted from the median is positive. Asymmetry can also be calculated from the skewness (Sokal and Rohlf, 1995). A normal distribution will have its skewness as 0, and a negatively skewed distribution will have negative skewness. Table 1 reports the mean, median, and skewness of each population of $\delta^{13}C_S$ presented in Figs. 7 and 8. When the mean, median, and skewness are calculated for the entire sample population, in general Model II (GMR and ODR) regression is negatively skewed, whereas Model I (OLS) is not (ODR results not shown in Table 1). What is interesting is that the mean, median, and skewness calculated from a Keeling GMR with simulated data with normally distributed random errors with 0.15 ppm and 0.15‰ standard deviations has its mean (−27.60‰) less than its median (−26.07‰), yet it has a skewness close to zero, suggesting a normal distribution. This may be due to the influence of outliers. Due to the large random errors on $\delta^{13}C$ (0.15‰), it was possible to obtain simulation values of $\delta^{13}C_S$ (>0 or <−50‰) that were not biologically plausible. This occurred when the [CO$_2$] range was small. To minimize this effect, we recalculated the mean, median, and skewness for $\delta^{13}C_S$ that fell within 10‰ of −25.67‰, the correct $\delta^{13}C_S$ with no error added. When this was done (right columns of Table 1) Model II regression consistently had negative skewness.

Fig. 10 shows fitted results of the theoretical uncertainty in $\delta^{13}C_S$ calculated from Eq. (13) versus [CO$_2$] range. This is similar to Fig. 1, but Fig. 10 contains the results of six different simulations. The uncertainty in [CO$_2$] was normally distributed with a standard deviation of 0.15 ppm for each simulation in the top panels of Fig. 10, and 0.10 ppm in the bottom panels. The uncertainty in $\delta^{13}C$ was normally distributed with a standard deviation of 0.15‰, 0.01‰, or 0.005‰. The top panels are representative of the analytical uncertainty in Bowling et al. (2005), and the lower panels are representative of the analytical uncertainty in Miller and Tans (2003). [CO$_2$] range and $\delta^{13}C_S$ uncertainty were then fit to a power function. ($y = ax^b$ where y is $\delta^{13}C_S$ error, x is [CO$_2$] range.) Fig. 10 shows the results for OLS and GMR. ODR was omitted because similar results to GMR were obtained. Note that in Fig. 10 there is no appreciable difference between Models I and II regression for the uncertainty in $\delta^{13}C_S$. However, note the significant reduction in $\delta^{13}C_S$ error at low [CO$_2$] ranges when the precision on $\delta^{13}C$ measurements is improved.

To investigate whether improvements in [CO$_2$] versus $\delta^{13}C$ measurement capability might decrease uncertainty in $\delta^{13}C_S$, we ran additional simulations with varying degrees of error in [CO$_2$] and $\delta^{13}C$ and then fitted $\delta^{13}C_S$ error versus [CO$_2$] range to a power function for each simulation. The motivation for this was to examine $\delta^{13}C_S$ uncertainty for typical night ranges (50 ppm) and typical day ranges (5 ppm). Corresponding isotope ratio variation in C$_3$ ecosystems is conservatively −0.05‰/ppm (Bowling et al., 1999), so these CO$_2$ ranges correspond to a $\delta^{13}C$ signal of 2.5‰ and 0.25‰, respectively. We then compared signal to
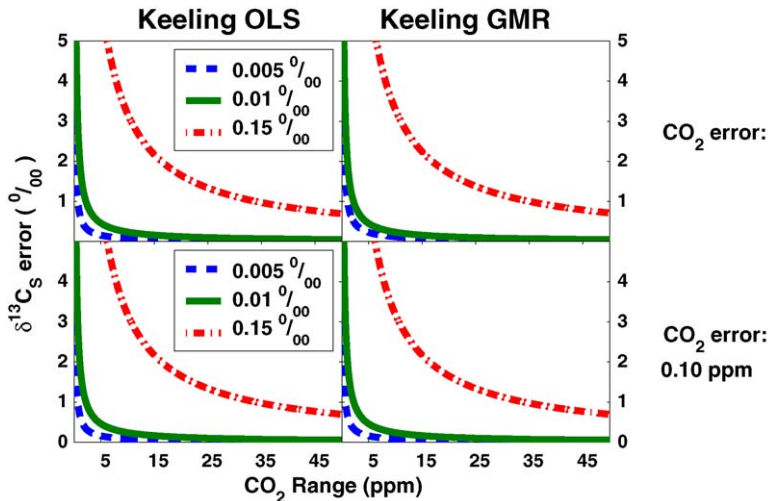
Fig. 10. Fitted results of the theoretical uncertainty in $\delta^{13}C_S$ calculated from Eq. (13) vs. [CO_2] range for Keeling OLS (left panels) and Keeling GMR (right panels). For ODR, similar results to GMR were obtained and hence omitted. The top panels come from simulations where [CO_2] have been randomly perturbed by 0.15 ppm, corresponding to the analytical uncertainty of [CO_2] in Bowling et al. (2005). The lower panels are from datasets where CO_2 has been randomly perturbed by 0.10 ppm, corresponding to the analytical uncertainty in Miller and Tans (2003). The legends correspond to the magnitude of the perturbation in $\delta^{13}C$.

noise ratios of [CO_2] and $\delta^{13}C$ and the resulting $\delta^{13}C_S$ error for each simulation. The results are shown in Table 2. The fifth and sixth columns of Table 2 highlight the fact that the signal to noise ratios of [CO_2] and $\delta^{13}C$ are not equal with present analytical capabilities. At current measurement capability of the TDL (precision of 0.15 ppm and 0.15‰), it is possible to determine $\delta^{13}C_S$ at an uncertainty of 0.71‰ when a sample [CO_2] range of 50 ppm is obtained. However, if a sample range of 5 ppm is obtained, the TDL can only determine $\delta^{13}C_S$ with an uncertainty near 6‰. This contrasts sharply with IRGA/mass spectrometry (Miller and Tans, 2003), which can resolve $\delta^{13}C_S$ to an uncertainty of 0.05‰ for a 50 ppm range, and 0.46‰ for a 5 ppm range of [CO_2]. Furthermore, the last two columns of Table 2 indicate that Model I and Model II regression give equivalent results for uncertainty in $\delta^{13}C_S$. If we wish to calculate $\delta^{13}C_S$ at small [CO_2] ranges – which is needed to interpret the daytime isotopic signals – improving the precision of [CO_2] measurements *will not* have as great an effect as improving the precision of $\delta^{13}C$ measurements because [CO_2] is currently measured at a relatively higher precision than $\delta^{13}C$.

The conclusion that improving analytical uncertainty on $\delta^{13}C$ more than [CO_2] will improve $\delta^{13}C_S$ precision can also be inferred from examining the error reduction factors (α and β in Eq. (13)). Table 3 lists the error reduction factor for [CO_2] and $\delta^{13}C$ for the TDL dataset. Note that the error reduction factor for $\delta^{13}C$ is three orders of magnitude larger than the reduction

factor for [CO_2], regardless of regression choice or isotopic mixing line.

We recognize the complexity associated with Eq. (13) to determine the error in $\delta^{13}C_S$ for a particular regression choice and isotopic mixing line. In Appendix A, we calculate the sensitivity factors α and β for a Miller/Tans OLS regression. Due to the equivalence in Fig. 10 between the $\delta^{13}C_S$ error for a Model I and a Model II regression, we can justify the use of Model I standard error to determine $\delta^{13}C_S$ error, even when Model II regression is used. Eq. (13) and the standard error of the OLS intercept (Keeling mixing line) or slope (Miller/Tans mixing line) give similar results for simulations (results not shown). For practical convenience, we recommend using these OLS standard error formulas as recommended by Pataki et al. (2003b) to determine uncertainty in $\delta^{13}C_S$. We do not expect this result to generalize to all studies, but this conclusion is a consequence of the fact that [CO_2] is measured with higher relative precision than $\delta^{13}C$. The fitted theoretical relationship in $\delta^{13}C_S$ uncertainty [CO_2] range for 0.15 ppm, 0.15‰ measurement uncertainty is shown in Fig. 1. Note that most of the uncertainty in $\delta^{13}C_S$ for the TDL fall at or below the fitted relationship.

To reduce the effect of the bias, one possible measurement protocol is to increase the number of samples. Due to the complexity of Eqs. (1)–(5), it is difficult to disentangle the effect of the range from the number of samples on parameter estimates. Fig. 5 shows results where the number of regression points is

Table 2
Calculation of signal to noise ratios for CO_2 and $\delta^{13}C$ with the corresponding uncertainty in $\delta^{13}C_S$ determined from fitting the theoretical error vs. CO_2 range to a power function $y = ax^b$ where $x = 50$ or $x = 5$

CO_2 signal (ppm)	$\delta^{13}C$ signal (‰)	Uncertainty in CO_2 (ppm)	Uncertainty in $\delta^{13}C$ (‰)	CO_2 SNR	$\delta^{13}C$ SNR	kOLS (‰)	kGMR (‰)
50	**2.50**	**0.15**	**0.15**	**333.3**	**16.67**	**0.71**	**0.71**
50	2.50	0.15	0.05	333.3	50	0.23	0.23
50	2.50	0.15	0.01	333.3	250	0.05	0.05
50	2.50	0.15	0.0075	333.3	333.3	0.04	0.04
50	2.50	0.15	0.0005	333.3	5000	0.03	0.03
50	2.50	0.1	0.15	500	16.67	0.69	0.69
50	2.50	0.1	0.05	500	50	0.23	0.23
50	**2.50**	**0.1**	**0.01**	**500**	**250**	**0.05**	**0.05**
50	2.50	0.1	0.0075	500	333.3	0.04	0.04
50	2.50	0.1	0.005	500	5000	0.02	0.02
50	2.50	0.01	0.15	5000	16.67	0.70	0.70
50	2.50	0.01	0.05	5000	50	0.23	0.23
50	2.50	0.01	0.01	5000	250	0.05	0.05
50	2.50	0.01	0.0075	5000	333.3	0.04	0.04
50	2.50	0.01	0.0005	5000	5000	0.003	0.003
50	2.50	0.005	0.15	10000	16.67	0.69	0.69
50	2.50	0.005	0.05	10000	50	0.23	0.23
50	2.50	0.005	0.01	10000	250	0.04	0.04
50	2.50	0.005	0.0075	10000	333.3	0.03	0.03
50	2.50	0.005	0.005	10000	5000	0.003	0.003
5	**0.25**	**0.15**	**0.15**	**33.3**	**1.67**	**6.36**	**6.36**
5	0.25	0.15	0.05	33.3	5	2.02	2.02
5	0.25	0.15	0.01	33.3	25	0.45	0.45
5	0.25	0.15	0.0075	33.3	33.3	0.40	0.39
5	0.25	0.15	0.0005	33.3	500	0.23	0.23
5	0.25	0.1	0.15	50	1.67	6.15	6.15
5	0.25	0.1	0.05	50	5	2.08	2.08
5	**0.25**	**0.1**	**0.01**	**50**	**25**	**0.46**	**0.46**
5	0.25	0.1	0.0075	50	33.3	0.36	0.36
5	0.25	0.1	0.0005	50	500	0.16	0.16
5	0.25	0.01	0.15	500	1.67	6.25	6.25
5	0.25	0.01	2.05	500	5	2.05	2.05
5	0.25	0.01	0.01	500	25	0.42	0.42
5	0.25	0.01	0.0075	500	33.3	0.32	0.32
5	0.25	0.01	0.0005	500	500	0.03	0.03
5	0.25	0.005	0.15	1000	1.67	6.30	6.30
5	0.25	0.005	0.05	1000	5	2.03	2.03
5	0.25	0.005	0.01	1000	25	0.39	0.39
5	0.25	0.005	0.0075	1000	33.3	0.31	0.03
5	0.25	0.005	0.005	1000	500	0.02	0.02

Values in bold represent current measurement capabilities with the TDL (0.15 ppm, 0.15‰) (Bowling et al., 2005) or IRGA/mass spectrometry (0.10 ppm, 0.01‰) (Miller and Tans, 2003). kOLS represents the uncertainty in $\delta^{13}C_S$ for a Keeling OLS regression, kGMR a $\delta^{13}C_S$ Keeling GMR regression uncertainty. For ODR, similar results to GMR were obtained and hence omitted.

increased from 4 to 16. For $n = 4$, there are more estimates of $\delta^{13}C_S$ at small ranges (which is to be expected) and by inspection the bias problem is less pronounced.

Clearly, increasing the number of samples can generate a dataset more representative of the mixing line. The number of samples is usually study dependent, but for mixing lines, 20 samples is a conservative estimate. All the regression estimates (OLS, GMR, and ODR) rely on minimizing some residual from the best-fit line (see Fig. 2). A sample of data points with similar x values (i.e. a small range) should have similar residuals from the best-fit line. Fig. 11 shows results of simulations where the number of samples was constrained to be either $n = 5$, 20, or 80. (Keeling OLS and GMR estimates only; similar results were obtained for Miller/Tans and ODR and hence omitted.) By inspection there is no apparent difference between the number of samples and the strength of Model II bias. Because of this discrepancy between simulated results

Table 3
Calculation of error reduction factors for OLS and GMR discussed in the text using TDL data from all measurement heights

Regression	CO_2 reduction factor, α	$\delta^{13}C$ reduction factor, β
kOLS	0.000295	0.203
mtOLS	0.000278	0.207
kGMR	0.000313	0.203
mtGMR	0.000277	0.208

For ODR, similar results to GMR were obtained and hence omitted. The initials preceding the regression types denote a Keeling (k) or a Miller/Tans (mt) isotopic mixing line.

(Fig. 11) and actual data (Fig. 5), we hypothesize that the differences between the estimates in Fig. 5 are due to biological and environmental variability. At small ranges with a low number of samples, it may be difficult to capture the linear relationship between 1/$[CO_2]$ and $\delta^{13}C$, and the mixing line is extremely sensitive to biological variation that generate large outliers from the best-fit line. At low ranges with a large number of samples, these outliers have less of an effect because the mixing line (albeit at small ranges) is well sampled for that range.

4.4. Recommendation of regression type and isotopic mixing line

One application of Keeling or Miller/Tans isotopic mixing lines is to compare estimates of $\delta^{13}C_S$ across biomes as done in Pataki et al. (2003b). As stated in Section 3.1, determining $\delta^{13}C_S$ from $[CO_2]$ and $\delta^{13}C$

measurements is an example of functional regression rather than forecasting values of $\delta^{13}C$ for a given $[CO_2]$ measurement. Model II regression has been traditionally recommended in this case because it removes the influence of determining an independent and dependent variable. Model II regression is heavily used in allometry studies where no ''independent'' variable exists (Henry and Aarssen, 1999). Pataki et al. (2003b) used this logic to advocate Model II regression for isotopic mixing lines.

At low $[CO_2]$ ranges, our analytical ability to measure a relationship between $[CO_2]$ and $\delta^{13}C$ is decreased. Model II regression is not appropriate when there is no apparent relationship in the data (Sprent and Dolby, 1980). If a relationship is still posited between the data, then Model I regression is recommended in order to get some idea (albeit with large errors) of the functional relationship (Sprent and Dolby, 1980). For isotopic mixing lines, mass conservation (Eqs. (7) and (8)) still applies at low $[CO_2]$ ranges, however it is difficult to determine $\delta^{13}C_S$ at these ranges due to a low signal to noise ratio in the measurements (Table 2). Fig. 4 shows that using OLS regression to estimate $\delta^{13}C_S$ will be normally distributed about the coefficients. Data from simulations were generated from a straight mixing line with $\delta^{13}C_S = -25.67‰$. When $[CO_2]$ and $\delta^{13}C_S$ data were perturbed with random error, simulations in Figs. 7 and 8 demonstrate that OLS regression estimates $\delta^{13}C_S$ in a range distributed about the ''correct'' $\delta^{13}C_S$ whereas Model II regression does not. Fig. 11 also demonstrates that there is no bias (plotted as $\delta^{13}C_S$ less the true value of $-25.67‰$)
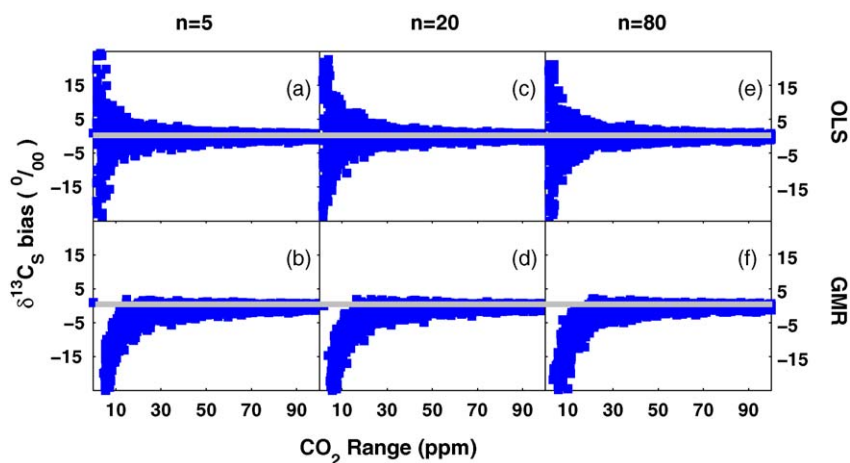

Fig. 11. Simulation results of $\delta^{13}C_S$ bias vs. $[CO_2]$ range for mixing lines with a variable number of samples ($n = 5$, 20 or 80). The gray line marks 0‰. The $\delta^{13}C_S$ bias is calculated as $\delta^{13}C_S$ less the value from a straight mixing line ($-25.67‰$). All the simulation results come from data that have been randomly perturbed by 0.15 ppm in $[CO_2]$, 0.15‰ in $\delta^{13}C$. The top panels represent Keeling OLS estimates of $\delta^{13}C_S$ and the bottom panels are Keeling GMR estimates of $\delta^{13}C_S$. Similar results for ODR regression techniques and Miller/Tans mixing lines were obtained and hence omitted.

associated with OLS across a range of sample sizes. This conclusion of using OLS at low ranges has been examined in other studies (Angleton and Bonham, 1995).

For large $[CO_2]$ ranges the recommendation to use GMR from Pataki et al. (2003b) gives reasonable results. However, as shown in Figs. 7 and 8, GMR regressions are unbiased compared to OLS only for large $[CO_2]$ and high r^2. This result may affect studies where GMR is used to estimate $\delta^{13}C_S$. If $\delta^{13}C_S$ was calculated using Model II regression from a mixing line with a low $[CO_2]$ range, we should expect a more negative calculation of $\delta^{13}C_S$. The amount of the bias is dependent on the level of analytical uncertainty in $[CO_2]$ and $\delta^{13}C$. Our sensitivity analysis demonstrates that reducing measurement uncertainty reduces our uncertainty in $\delta^{13}C_S$, and hence will reduce the effect of the bias (Figs. 7 and 8).

Pataki et al. (2003b) argue that Model II regression should be used over Model I because there is no clear independent variable in isotopic mixing lines. Because of comparatively greater signal to noise ratios in $\delta^{13}C$ measurements than $[CO_2]$ measurements (Table 2), and the lack of a bias in Model I regression in real data (Figs. 3 and 4) and simulated data (Figs. 6–8 and 11) the use of Model I regression with $\delta^{13}C$ as the dependent variable is justifiable for isotopic mixing lines.

As shown in Section 3.2, the Keeling and Miller/Tans isotopic mixing lines are derived from the same conservation equations. In theory they should give an equivalent estimate of $\delta^{13}C_S$. Simulations shown in Figs. 7 and 8 show no inherent difference between the two, and the theoretical uncertainty analysis for $\delta^{13}C_S$ does not illustrate an advantage of one over the other three. Note, however, the Miller/Tans formulation of the isotopic mixing line offers an advantage when the background values can be specified (Miller and Tans, 2003; Lai et al., 2004, 2005).

Differences were found between $\delta^{13}C_S$ calculated from a Keeling or a Miller/Tans regression for Model II regression. The top panel of Fig. 12 demonstrates that the difference between the two (γ) increases as $[CO_2]$ range decreases for *Model II only*, whereas the difference for Model I does not. Using a Model II regression to reduce the influence of the independent variable gives rise to a bias as discussed because Keeling and Miller/Tans regressions occur in different dimensional space. GMR and ODR have different strategies to minimize the vertical and horizontal residuals. Conceptually shown in Fig. 2, GMR performs two OLS regressions and takes the geometric mean of them. ODR minimizes the perpendicular residual from
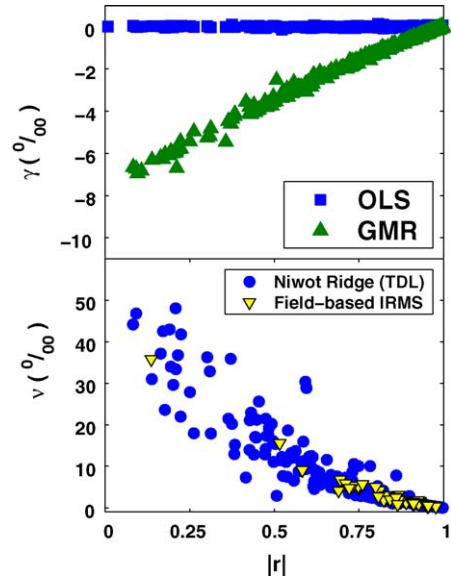

Fig. 12. (Top panel) Comparison of the difference between $\delta^{13}C_S$ calculated from a Keeling mixing line and a Miller/Tans mixing line for OLS and GMR using the TDL data (nighttime and daytime). On the vertical axis $\gamma = \delta^{13}C_{S_{Keeling}} - \delta^{13}C_{S_{Miller/Tans}}$. For ODR, similar results to GMR were obtained and hence omitted. Bottom panel: Comparison of the difference between $\delta^{13}C_S$ calculated from a Miller/Tans mixing line using OLS and a Miller/Tans mixing line using GMR for the Niwot Ridge TDL (Bowling et al., 2005) and the 45 min Grünschwaige field-based IRMS (Schnyder et al., 2004). On the vertical axis $\nu = \delta^{13}C_{S_{OLS}} - \delta^{13}C_{S_{GMR}}$.

the hypothetical best-fit line. Because of these different strategies, GMR and ODR produce a bias that is dependent on the correlation between the variables. Because there is no bias associated with Model I regression, this is further justification for its use in this application.

An implication of this study is the re-interpretation of highly negative values of $\delta^{13}C_S$ in studies that use geometric mean regression (Bowling et al., 2002; Pataki et al., 2003a; McDowell et al., 2004; Lai et al., 2004, 2005; Hemming et al., 2005). Highly negative values of $\delta^{13}C_S$ may be indicative of anthropogenic effects (Pataki et al., 2003a; Lai et al., 2004) or situations when there is a small isotopic disequilibrium between a source and a sink of $[CO_2]$ (this may occur with photosynthesis and respiration, see Miller and Tans (2003)). If $\delta^{13}C_S$ was calculated from a mixing line with low r, then the result will be even more biased. The bottom panel of Fig. 12 presents the difference (ν) between $\delta^{13}C_S$ calculated via a Miller/Tans OLS regression and a Miller/Tans GMR regression using the TDL data, separated into upper canopy inlets (5, 7, 9, and 11 m aboveground) and lower canopy inlets (0.1, 0.5, 1 and 2 m aboveground) as done in Bowling et al.

(2005) and data from the field-based mass spectrometer at Grünschwaige. The lower half of Fig. 11 clearly demonstrates low r may bias $\delta^{13}C_S$ by 50‰. The magnitude of the bias is dependent on the analytical uncertainty of the measurements.

Fig. 6 demonstrates the marked contrast between the BASIN dataset and datasets collected by TDL and field-based mass spectrometry. As described in Section 4.1, this difference may be attributed to higher precision on the BASIN dataset or rejection of data points of poor quality. The Niwot Ridge and Grünschwaige studies have a higher proportion of poor quality datasets (low r) at low $[CO_2]$ range; this is also reflected in simulated data as well at all levels of error in $[CO_2]$ and $\delta^{13}C$. We hypothesize the difference between the two is rejection of data of poor quality (either outliers or datasets with low $[CO_2]$ range). If data are also excluded because they seem "biologically meaningless," an implication is biased (prejudiced) data reporting of $\delta^{13}C_S$ values. To resolve this epistemological issue on the criteria for "good" data, we advocate a thorough discussion about objectivity in data elimination and reporting.

The mathematical results of this study (increased Model II bias at low r) and sensitivity analysis (Eq. (13)) can apply to many diverse areas outside of carbon cycle research (e.g. isotopic diet studies, Best and Schell, 1996; Burton and Koch, 1999; tree allometry studies, Henry and Aarssen, 1999; or ecotoxicology studies, Angleton and Bonham, 1995). To obtain consistent parameter estimates via linear regression, it is necessary to understand (a) the differences between different types of linear regression (OLS, GMR, and ODR), and (b) the influence of regression choice on the calculation and interpretation of regression parameters.

5. Conclusions

From this study we conclude the following:

- The use of Model II regression to determine $\delta^{13}C_S$ is inappropriate because it is a biased estimator of $\delta^{13}C_S$. On the other hand, Model I regression gives *unbiased* estimates of $\delta^{13}C_S$ at all $[CO_2]$ ranges with all relevant scenarios of instrument precision (encompassing laboratory-based dual inlet mass spectrometry field-based mass spectrometry, and TDL). This conclusion is also justified from simulations of isotopic mixing lines from $[CO_2]$ and $\delta^{13}C$ data that have varying levels of random error.
- The standard error formulas for Model I regression are fine as a measure of $\delta^{13}C_S$ uncertainty.

- There is no inherent advantage or disadvantage to using either a Keeling or a Miller/Tans approach to determine $\delta^{13}C_S$.
- Our ability to estimate $\delta^{13}C_S$ at low $[CO_2]$ ranges is determined primarily on the measurement uncertainties of $\delta^{13}C$. Instrument development is necessary to minimize this uncertainty in order to minimize $\delta^{13}C_S$ uncertainty.

Appendix A

Assume that we have N measurements of $[CO_2]$ and $\delta^{13}C$. Denote the ith measurement by; $[CO_2]$ by c_i and $\delta^{13}C$ by δ_i. The slope of a Miller/Tans OLS regression is given by

$$b = \frac{N\sum c_j^2\delta_j - \sum c_j\sum c_j\delta_j}{N\sum c_j^2 - \left(\sum c_j\right)^2}. \tag{18}$$

The error reduction factor for $[CO_2]$ is given by

$$\frac{\partial b}{\partial c_i} =$$

$$\frac{\left(2Nc_i\delta_i - \sum c_j\delta_j - \delta_i\sum c_j\right)\left(N\sum c_j^2 - \left(\sum c_j\right)^2\right)}{\left(N\sum c_j^2 - \left(\sum c_j\right)^2\right)^2}$$
$$\frac{- \left(N\sum c_j^2\delta_j - \sum c_j\sum c_j\delta_j\right)\left(2Nc_i - 2\sum c_j\right)}{\left(N\sum c_j^2 - \left(\sum c_j\right)^2\right)^2}$$

$$= \frac{2c_i\delta_i - \overline{\delta_j c_j} - \delta_i\bar{c}_j - 2b(c_i - \bar{c}_j)}{\sum c_j^2 - N^{-1}\left(\sum c_j\right)^2}, \tag{19}$$

where $\bar{x}_j$ represents the mean of the variable x. Similarly, the error reduction for $\delta^{13}C$ is given by

$$\frac{\partial b}{\partial \delta_i} = \frac{Nc_i^2 - c_i \sum c_j}{N \sum c_j^2 - \left(\sum c_j\right)^2} = \frac{c_i^2 - \bar{c}_j c_i}{N \sum c_j^2 - \left(\sum c_j\right)^2}. \tag{20}$$

References

Angleton, G.M., Bonham, C.D., 1995. Least squares regression vs. geometric mean regression for ecotoxicology studies. Appl. Math. Comput. 72, 21–32.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: a new tool to study the temporal and dioxide, water vapor, and energy flux. Am. Meterol. Soc. 82 (11), 2415–2434.

Best, P.B., Schell, D.M., 1996. Stable isotopes in southern right whale (*Eubalaena australis*) baleen as indicators of seasonal movements, feeding and growth. Mar. Biol. 124, 483–494.

Boggs, P.T., Rogers, J.E., 2001. Orthogonal distance regression. In: Brown, P.J., Fullers, W.A. (Eds.), Statistical Analysis of Measurement Error Models and their Applications, vol. 112 of Contemporary Mathematics. American Mathematical Society, Providence, RI, pp. 183–194.

Bowling, D.R., Baldocchi, D.D., Monson, R.K., 1999. Dynamics of isotopic exchange of carbon dioxide in a Tennessee deciduous forest. Global Biogeochem. Cycles 13, 903–922.

Bowling, D.R., Burns, S.P., Conway, T.J., Monson, R.K., White, J.W.C., 2005. Extensive observations of CO_2 carbon isotope content in and above a high-elevation subalpine forest. Global Biogeochem. Cycles 19 (3), doi:10.1029/2004GB002394.

Bowling, D.R., McDowell, N.G., Bond, B.J., Law, B.E., Ehleringer, J.R., 2002. ^{13}C content of ecosystem respiration is linked to precipitation and vapor pressure deficit. Oecologia 131, 113–124.

Bowling, D.R., McDowell, N.G., Welker, J.M., Bond, B.J., Law, B.E., Ehleringer, J.R., 2003a. Oxygen isotope content of CO_2 in nocturnal ecosystem respiration. 1. Observations in forests along a precipitation transect in Oregon, USA. Global Biogeochem. Cycles 17 (4), 1120, doi:10.1029/2003GB002081.

Bowling, D.R., McDowell, N.G., Welker, J.M., Bond, B.J., Law, B.E., Ehleringer, J.R., 2003b. Oxygen isotope content of CO_2 in nocturnal ecosystem respiration. 2. Short-term dynamics of foliar and soil component fluxes in an old-growth ponderosa pine forest. Global Biogeochem. Cycles 17 (4), 1124, doi:10.1029/2003GB002082.

Buchmann, N., Cuehl, J.M., Barigah, T.S., Ehleringer, J.R., 1997a. Interseasonal comparison of CO_2 concentrations, isotopic composition, and carbon dynamics in an Amazonian rainforest (French Guiana). Oecologia 110, 120–131.

Buchmann, N., Kao, W.Y., Ehleringer, J.R., 1997b. Influence of stand structure on carbon-13 of vegetation, soils, and canopy air within deciduous and evergreen forests in Utah, United States. Oecologia 110, 109–119.

Burton, R.K., Koch, P.L., 1999. Isotopic tracking of foraging and long-distance migration in northeastern pacific pinnipeds. Oecologia 119, 578–585.

Flanagan, L.B., Brooks, J.R., Varney, G.T., Berry, S.C., Ehleringer, J.R., 1996. Carbon isotope discrimination during photosynthesis and the isotope ratio of respired CO_2 in boreal forest ecosystems. Global Biogeochem. Cycles 10, 629–640.

Goulden, M.L., Munger, J.W., Fan, S.-M., Daube, B.C., Wofsy, S.C., 1996. Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. Science 271, 169–182.

Gubbins, D., 2004. Time Series Analysis and Inverse Theory for Geophysicists. Cambridge University Press, Cambridge, UK.

Hemming, D., Yakir, D., Ambus, P., Aurela, M., Besson, C., Black, K., Buchmann, N., Burlett, R., Cescatti, A., Clement, R., Gross, P., Granier, A., Grunwald, T., Havrankova, K., Janous, D., Janssens, I.A., Knohl, A., Kostner, B., Kowalski, A., Laurila, T., Mata, C., Marcolla, B., Matteucci, G., Moncrieff, J., Moors, E.J., Osborne, B., Santos Periera, J., Pihlatie, M., Pilegaard, K., Ponti, F., Rosova, Z., Rossi, F., Scartazza, A., Vesala, T., 2005. Pan-European $\delta^{13}C$ values of air and organic matter from forest ecosystems. Global Change Biol. 11, 1065–1093.

Henry, H.A.L., Aarssen, L.W., 1999. The interpretation of stem diameter–height allometry in trees: biomechanical constraints, neighbour effects, or biased regressions? Ecol. Lett. 2, 89–97.

Huxman, T.E., Turnipseed, A.A., Sparks, J.P., Harley, P.C., Monson, R.K., 2003. Temperature as a control over ecosystem CO_2 fluxes in a high-elevation, subalpine forest. Oecologia 134, 537–546.

Janssens, I.A., Lankreijer, H., Matteucci, G., Kowalski, A.S., Buchmann, N., Epron, D., Pilegaard, K., Kutsch, W., Longdoz, B., Gruenwald, T., Montagnani, L., Dore, S., Rebmann, C., Moors, E.J., Grelle, A., Rannik, U., Morgenstern, K., Oltchev, S., Clement, R., Gudmundsson, J., Minerbi, S., Berbigier, P., Ibrom, A., Moncrieff, J., Aubinet, M., Bernhofer, C., Jensen, N.O., Vesala, T., Granier, A., Schulze, E.-D., Lindroth, A., Dolman, A.J., Jarvis, P.G., Ceulemans, R., Valentini, R., 2001. Productivity overshadows temperature in determining soil and ecosystem respiration across European forests. Global Change Biol. 7, 269–278.

Keeling, C., 1958. The concentrations and isotopic abundances of atmospheric carbon dioxide in rural areas. Geochim. Cosmochim. Acta 13, 322–344.

Knohl, A., Werner, R.A., Brand, W.A., Buchmann, N., 2005. Short-term variations in $\delta^{13}C$ of ecosystem respiration reveals link between assimilation and respiration in a deciduous forest. Oecologia 140, 70–82.

Lai, C.T., Ehleringer, J.R., Schauer, A.J., Tans, P.P., Hollinger, D.Y., Paw, U.K.T., Munger, J.W., Wofsy, S.C., 2005. Canopy-scale $\delta^{13}C$ of photosynthetic and respiratory CO_2 fluxes: observations in forest biomes across the United States. Global Change Biol. 11, 633–643.

Lai, C.T., Ehleringer, J.R., Tans, P.P., Wofsy, S.C., Urbanski, S.P., Hollinger, D.Y., 2004. Estimating photosynthetic ^{13}C discrimination in terrestrial CO_2 exchange from canopy to regional scales. Global Biogeochem. Cycles 18, doi:10.1029/2003GB002148.

Laws, E., 1997. Mathematical Methods for Oceanographers. John Wiley, New York.

McDowell, N.G., Bowling, D.R., Schauer, A., Irvine, J., Bond, B.J., Law, B.E., Ehleringer, J.R., 2004. Associations between carbon isotope ratios of ecosystem respiration, water availability and canopy conductance. Global Change Biol. 10, 1767–1784.

Miller, J.B., Tans, P.P., 2003. Calculating isotopic fractionation from atmospheric measurements at various scales. Tellus 55B, 207–214.

Miller, J.B., Tans, P.P., White, J.W.C., Conway, T.J., Vaughn, B.W., 2003. The atmospheric signal of terrestrial carbon isotopic discrimination and its implication for partitioning carbon fluxes. Tellus 55B, 197–206.

Pataki, D.E., Bowling, D.R., Ehleringer, J.R., 2003a. Seasonal cycle of carbon dioxide and its isotopic composition in an urban atmosphere: anthropogenic and biogenic effects. J. Geophys. Res. 108, doi:10.1029/2003JD003865.

Pataki, D.E., Ehleringer, J.R., Flanagan, L.B.D.Y., Bowling, D.R., Still, C.J., Buchmann, N., Kaplan, J.O., Berry, J., 2003b. The application and interpretation of Keeling plots in terrestrial carbon cycle research. Global Biogeochem. Cycles 17 (1), 1022, doi:10.1029/2001GB001850.

Piovesan, G., Adams, J.M., 2000. Carbon balance gradient in European forests: interpreting EUROFLUX. J. Veg. Sci. 11, 923–926.

Prentice, I.C., Farquhar, G.D., Fasham, M.J.R., Goulden, M.L., Heimann, M., Jaramillo, V.J., Kheshgi, H.S., Le Quéré, C., Scholes, R.J., Wallace, D.W.R., Archer, D., Ashmore, M.R., Aumont, O., Baker, D., Battle, M., Bender, M., Bopp, L.P., Bousquet, P., Caldeira, K., Ciais, P., Cox, P.M., Cramer, W., Dentener, F., Enting, I.G., Field, C.B., Friedlingstein, P., Holland, E.A., Houghton, R.A., House, J.I., Ishida, A., Jain, A.K., Janssens, I.A., Joos, F., Kaminski, T., Keeling, C.D., Keeling, R.F., Kicklighter, D.W., Kohfeld, K.E., Knorr, W., Law, R., Lenton, T., Lindsay, K., Maier-Reimer, E., Manning, A.C., Matear, R.J., McGuire, A.D., Melillo, J.M., Meyer, R., Mund, M., Orr, J.C., Piper, S., Plattner, K., Rayner, P.J., Sitch, S., Slater, R., Taguchi, S., Tans, P.P., Tian, H.Q., Weirig, M.F., Whorf, T., Yool, A., Pitelka, L., Ramirez Rojas, A., 2001. The carbon cycle and atmospheric carbon dioxide. In: Houghton, J.T., Yihui, D. (Eds.), Climate Change 2001: The Scientific Basis, vol. 11Cambridge University Press, pp. 183–237 (Chapter 3).

Ricker, W.E., 1973. Linear regression in fishery research. J. Fish. Res. Board Canada 30, 409–434.

Scartazza, A., Mata, C., Matteucci, G., Yakir, D., Moscatello, S., Brugnoli, E., 2004. Comparisons of $\delta^{13}C$ of photosynthetic products and ecosystem respiratory CO_2 and their responses to seasonal climate variability. Oecologia 140, 340–351.

Schimel, D.S., Brawell, B.H., Holland, E.A., McKeown, R., Ojima, D.S., Painter, T.H., Parton, W., Townsend, A.R., 1994. Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. Global Biogeochem. Cycles 8, 279–293.

Schimel, D.S., House, J.I., Hibbard, K.A., Bousquet, P., Ciais, P., Peylin, P., Braswell, B.H., Apps, M.J., Baker, D., Bondeau, A., Canadell, J., Churkina, G., Cramer, W., Denning, A.S., Field, C.B., Friedlingstein, P., Goodale, C., Heimann, M., Houghton, R.A., Melillo, J.M., Moore III, B., Murdiyarso, D., Noble, I., Pacala, S.W., Prentice, I.C., Raupach, M.R., Rayner, P.J., Scholes, R.J., Steffen, W.L., Wirth, C., 2001. Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. Nature 414, 169–172.

Schnyder, H., Schäufele, R., Wenzel, R., 2004. Mobile, outdoor continuous-flow isotope-ratio mass spectrometer system for automated high-frequency ^{13}C– and ^{18}O–CO_2 analysis for Keeling plot applications. Rapid Commun. Mass Spectrom. 18, 3068–3074.

Sokal, R., Rohlf, J., 1995. Biometry. W.H. Freeman and Company, New York.

Sprent, P., Dolby, G.R., 1980. The geometric mean functional relationship. Biometrics 36, 547–550.

Still, C.J., Berry, J.A., Ribas-Carbo, M., Helliker, B.R., 2003. The contribution of C_3 and C_4 plants to the carbon cycle of a tallgrass prairie: an isotopic approach. Oecologia 136, 347–359.

Tarantola, A., 2005. Inverse Problem Theory and Model Parameter Estimation. SIAM Books, Philadelphia, PA.

Taylor, J.R., 1997. An Introduction to Error Analysis, 2nd ed. University Science Books, Sausalito, CA.

Valentini, R., Matteucci, G., Dolman, A.J., Schulze, E.-D., Rebmann, C., Moors, E.J., Granier, A., Gross, P., Jensen, N.O., Pilegaard, K., Lindroth, A., Grelle, A., Bernhofer, C., Grünwald, T., Aubinet, M., Ceulemans, R., Kowalski, A.S., Vesala, T., Rannik, Ü., Berbigier, P., Loustau, D., Gudmundsson, J., Thorgeirsson, H., Ibrom, A., Morgenstern, K., Clement, R., Moncrieff, J., Montagnani, L., Minerbi, S., Jarvis, P.G., 2000. Respiration as the main determinant of carbon balance in European forests. Nature 404, 861–865.

Van Dijk, A.I.J.M., Dolman, A.J., 2004. Estimates of CO_2 uptake and release among European forests based on eddy covariance data. Global Change Biol. 10, 1445–1459.