Quantifying and reducing uncertainties in estimated soil CO_2 fluxes with hierarchical data-model integration

Kiona Ogle[1†], Edmund Ryan[2], Feike A. Dijkstra[3], Elise Pendall[4]

[1]School of Informatics, Computing & Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA

[2]Lancaster Environment Center, Lancaster University, Lancaster, UK

[3]Centre for Carbon, Water & Food, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

[4]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, Australia

*Corresponding author: Kiona Ogle (Kiona.Ogle@nau.edu)
†Current Address:
School of Informatics, Computing & Cyber Systems (SICCS)
Northern Arizona University
P.O. Box 5693
Flagstaff, AZ 86011
tele: (928) 523-6200
fax: (928) 523-1902 (c/o Ogle, SICCS)

Key points

- A hierarchical modeling approach facilitates accurate estimates of soil CO_2 fluxes
- The approach is applied to data obtained from non-steady state soil chambers
- A hierarchical, non-steady state diffusion model of CO_2 fluxes performed best

Abstract

Non-steady state chambers are often employed to measure soil CO_2 fluxes. CO_2 concentrations (*C*) in the headspace are sampled at different times (*t*), and fluxes (*f*) are calculated from regressions of *C* versus *t* based a limited number of observations. Variability in the data can lead to poor fits and unreliable *f* estimates; groups with too few observations or poor fits are often discarded, resulting in "missing" *f* values. We solve these problems by fitting linear (steady state) and non-linear (non-steady state, diffusion based) models of *C* versus *t*, within in a hierarchical Bayesian framework. Data are from the Prairie Heating and CO_2 Enrichment (PHACE) study that manipulated atmospheric CO_2, temperature, soil moisture, and vegetation. CO_2 was collected from static chambers bi-weekly during five growing seasons, resulting in >12,000 samples and >3100 groups and associated fluxes. We compare *f* estimates based on non-hierarchical and hierarchical Bayesian (B vs HB) versions of the linear and diffusion-based (L vs D) models, resulting in four different models (BL, BD, HBL, HBD). Three models fit the data exceptionally well ($R^2 \geq 0.98$), but the BD model was inferior ($R^2 = 0.87$). The non-hierarchical models (BL, BD) produced highly uncertain *f* estimates *f* (wide 95% CIs), whereas the hierarchical models (HBL, HBD) produced very precise estimates. Of the hierarchical versions, the linear model (HBL) underestimated *f* by ~33% relative to the non-steady state model (HBD). The hierarchical models offer improvements upon traditional non-hierarchical approaches to estimating *f*, and we provide example code for the models.

Index terms: (1) 0490, (2) 0414, (3) 1986, (4) 0428, (5) 1990

Key words: Bayesian modeling, borrowing of strength, diffusion equation, Fick's law, global change experiment, soil respiration

1. Introduction

Soils are primary sources or sinks of radiatively active "greenhouse" gases such as carbon dioxide (CO_2), and quantifying CO_2 fluxes has been the subject of intense research for the last few decades [e.g., *Raich and Schlesinger*, 1992]. CO_2 and other trace gas fluxes are typically measured by inserting a small chamber into or on top of the soil, and collecting gas samples at predetermined time intervals after closure to follow the change in concentration in the chamber headspace as the gas accumulates or is drawn-down due to soil production or consumption, respectively. The gas concentrations may be analyzed in the field, such as by an in-line infrared gas analyzer (IRGA) [e.g., *Davidson et al.*, 2002], or brought back to the lab and analyzed via an IRGA or gas chromatography (GC) [e.g., *Venterea et al.*, 2009]. If the gas concentration (C) changes approximately linearly with time (t) since closure, then the trace gas fluxes are typically estimated from linear, or sometimes non-linear, regressions of C versus t for each independent chamber session.

The typical regression approach, however, potentially suffers from three primary issues. First, CO_2 concentrations collected while the chamber is closed may deviate from linearity due to time-dependent feedbacks between soil air and chamber headspace [*Livingston et al.*, 2005]. For instance, such feedbacks can reduce the diffusion gradient as CO_2 builds up in the chamber and diffuses out laterally, leading to underestimation of CO_2 fluxes by up to 25% [*Livingston et al.*, 2005]. This problem can be addressed by fitting a non-linear model to the C versus t data, such as an exponential decay function [*Hutchinson and Mosier*, 1981], quadratic function [*Wagner et al.*, 1997] or, less commonly, models inspired by diffusion theory [*Livingston et al.*, 2006;

Pedersen et al., 2001]. Second, missing or highly variable observations can lead to poor regression fits (i.e., low R^2 value) for particular chamber sessions, for both linear and non-linear models. This problem can be addressed by collecting more data points in each chamber session [e.g., *Davidson et al.*, 2002], by grouping similar chamber sessions, or by discarding data for problematic chamber sessions [*Hart*, 2006; *Pihlatie et al.*, 2007]. Third, uncertainty estimates associated with each flux value are typically ignored, or if reported, they still are not accounted for in subsequent analysis or modeling of the flux estimates, which are treated like data. This issue can be addressed using statistical methods that quantify precision and propagate uncertainty such as Monte Carlo analysis [*Venterea et al.*, 2009], but such approaches are rarely utilized.

We overcome these three issues by developing a hierarchical Bayesian approach coupled with a non-linear, non-steady state flux model that is derived from fundamental diffusion theory [*Livingston et al.*, 2006]. We demonstrate how the hierarchical approach addresses the missing or "bad" data problem, propagates uncertainties in the individual flux estimates, and can easily accommodate a diffusion-based model to account for non-steady state conditions. We illustrate our modeling approach by applying it to data on C versus t that were obtained from the Prairie Heating and CO_2 Enrichment (PHACE) study conducted in a semiarid grassland in Wyoming. PHACE was a global change experiment involving manipulations of atmospheric [CO_2], temperature, soil moisture, and vegetation status, resulting in 12 different treatment combinations, with five plots (replicates) per treatment level. We focus on the CO_2 data to illustrate our modeling approach because it is an important greenhouse gas, and understanding controls on soil respiration is paramount to understanding the global carbon cycle [*Bond-Lamberty and Thomson*, 2010]. Moreover, because the soil acts as a source of CO_2 (C

accumulates in the chamber), we can draw upon existing, concise analytical solutions to the standard diffusion equation [*Livingston et al.*, 2006].

The objective of this study is to describe and illustrate a more robust method for estimating CO_2 fluxes from data generated from static chambers. First, we draw-upon on a non-steady state flux model that explicitly accounts for time dependent artifacts such as soil-chamber feedbacks [*Davidson et al.*, 2002; *Livingston et al.*, 2006]. Second, we employ a hierarchical statistical model that accommodates the nested and crossed design of the PHACE experiment by assuming that the session-level flux terms (parameters in the linear and non-steady state models) vary around treatment by sampling date fluxes. The hierarchical approach results in "borrowing of strength" or "partial pooling" [*Gelman and Hill*, 2007; *Gelman et al.*, 2012] among chamber sessions such that sessions associated with problematic data are informed by sessions with clean data. The Bayesian framework allows the uncertainty in the flux estimates to be easily propagated to subsequent analyses, which can be simultaneously implemented within the Bayesian flux model; we illustrate this by conducting a simple post-analysis to evaluate the effects of the global change treatments on soil CO_2 fluxes.

2. Field Methods

2.1. Field Experiment

Data were obtained as part of the Prairie Heating And CO_2 Enrichment (PHACE) experiment that was conducted in a semiarid mixed prairie in southeastern Wyoming, USA (41° 11' N, 104° 54' W). The vegetation is dominated by a mixture of C4 and C3 grasses, including *Bouteloua gracilis* (C4), *Pascopyrum smithii* (C3), and *Hesperostipa comata* (C3). The soil is a fine-loamy, mixed, mesic Aridic Argiustoll. The mean monthly air temperatures range from -2.5 °C in January to 17.5 °C in July, and the mean annual precipitation is 384 mm (based on 132 years of weather records). Chamber CO_2 data were collected during the growing seasons (April – October) of 2007 through 2011 (five years). The average air temperature during these growing seasons ranged from 12.5 °C (2009) to 17.4 °C (2007), and the total precipitation received during each growing season ranged from 300 mm (2010) to 425 mm (2009). The site conditions and climate during the study period are described in greater detail in Dijkstra et al. [2013] and Zelikova et al. [2015].

The PHACE study was established in 2005, at which time, 20 plots (3.4 m diameter) were assigned to one of four treatment combinations (5 plots per treatment): ambient CO_2 and temperature (denoted ct), ambient CO_2 and elevated temperature (cT), elevated CO_2 and ambient temperature (Ct), and both elevated CO_2 and temperature (CT). Free Air CO_2 Enrichment technology was used to raise the atmospheric [CO_2] to ~600 ppm (±40 ppm) in the elevated CO_2 plots (Ct and CT). Ceramic infrared heaters were used to raise the canopy temperature by about 1.5 °C and 3 °C above the ambient temperature during the day and night, respectively, in the elevated temperature plots (cT and CT). The CO_2 and warming treatments were initiated in April 2006 and April 2007, respectively. An additional 10 plots were established in 2007 and assigned to one of two irrigation treatments that experienced ambient CO_2 and temperatures (5 plots each): shallow irrigation (cts, 3-5 irrigation events during the growing season to maintain soil water content similar to that in elevated CO_2 plots) or deep irrigation (ctd, two irrigation events at the start and end of the growing season, annual amount equal to that in cts treatment). Additional details about the PHACE experiment and associated treatment methodologies are provided in Dijkstra et al. [2010] and LeCain et al. [2015].

In 2008, a 0.4 m^2 subplot was established in each of the ct, cT, Ct, and CT plots. The subplots were isolated from the surrounding plot by a metal sheet that was buried 30 cm into the soil, and vegetation in the subplots was killed by application of a broad spectrum systematic herbicide (glyphosate). Seedlings that emerged after herbicide application were manually removed. See Dijkstra et al. [2013] for details about the herbicide treatment.

2.2. Chamber CO_2 Measurements

We used static, closed chambers [*Hutchinson and Mosier*, 1981] to measure CO_2 fluxes approximately every other week during the growing season, resulting in 12-16 measurements each year, for five years (2007-2011). In each plot, chamber anchors (diameter 20 cm, height 10 cm) were inserted 8 cm into the soil one month prior to the first measurements. One anchor was placed in the area with intact vegetation, and one anchor in the subplots where vegetation was removed. Measurements were taken between 10:00 am and 1:00 pm local time, separated into three periods, with each period lasting one hour to measure 10 plots simultaneously. Treatments were blocked within each period to minimize biases caused by diurnal effects on trace gas fluxes.

Chambers were placed on the anchors and sealed with a rubber band (made from a tire inner tube). Headspace gas samples (20 mL) were taken immediately after placing the chambers on the anchors (time $t = 0$), and after $t = 15$, 30, and 45 minutes (for the first three measurements dates in 2007, gas samples were not taken at 45 minutes) and injected into 12 mL evacuated Exetainers (Labco Limited, Lampeter, UK). Gas samples were analyzed for CO_2 on a gas chromatograph (Varian 3800, Varian, Inc., Palo Alto, CA, USA) usually within two days after sampling (CO_2 was measured with a thermal conductivity detector). The minimum detection limit for CO_2 calculated according to Parkin and Venterea [2010] was 0.1 mg CO_2-C m^{-2} hr^{-1}. Data were available for 3139 chamber sessions, yielding 12,240 pairs of (C, t) observations.

2.3. Environmental Data

Continuous, plot-level measurements of soil temperature and water content were made throughout the PHACE experiment. Custom-built Type T thermocouples were used to monitor soil temperature at a depth of 3 cm within ~1 m of each chamber and logged on an hourly basis on a Campbell CR-1000 data loggers (Campbell Scientific, Logan, UT, USA); soil temperatures recorded at the time of each chamber session were used for this study. Volumetric soil water content was monitored in each plot at multiple depths using EnviroSMART sensors (Sentek Sensor Technologies, Stepney, Australia); for this study, we used the 5-15 cm data. Soil water data were missing for ca. 6% of the days, primarily due to instrument failure. We gap-filled missing values using data from a nearby plot belonging to the same experimental treatment, or using cubic spline interpolation on days when data were missing across all or most plots of the same treatment [see, *Ryan et al.*, 2015]. In this study, we used daily averages of the hourly soil water content values.

3. Estimating Soil CO_2 Fluxes

We evaluated two different process models and two different statistical modeling approaches to estimating soil CO_2 fluxes based on the aforementioned data (§2.2 and §2.3). One process model is based on a simple linear model of C versus t, and the other represents a non-linear, non-steady state model. For the statistical approaches, we fit the process models to all data in a non-hierarchical framework that treats each chamber session as an independent data set (akin to traditional approaches). We also fit the models to the data in a hierarchical statistical framework that views the chamber sessions as samples from a population of sessions, thus allowing for borrowing of strength [*Gelman et al.*, 2012] among chamber sessions. We begin with a description of the process models (linear followed by the non-steady state diffusion model), then we describe the statistical (non-hierarchical followed by hierarchical) approaches to fitting the process models to the chamber C and t data. All four model combinations are implemented in a Bayesian framework, which we will refer to as the BL (non-hierarchical Bayesian linear), HBL (hierarchical Bayesian linear), BD (non-hierarchical Bayesian non-steady state diffusion), and HBD (hierarchical Bayesian, non-steady state diffusion) models.

3.1. Linear Model

This model assumes a linear relationship between CO_2 concentration (C, μmol m^{-3}) and time since chamber closure (t, sec):

$$C_t = C_0 + f\frac{A}{V}t \tag{1.1}$$

where C_0 (μmol m^{-3}) is the initial [CO_2] in the chamber at time $t = 0$; f (μmol m^{-2} sec^{-1}) is the flux density across the soil-atmosphere interface at time $t = 0$; A (m^2) is the soil surface area over which the chamber is deployed; V (m^3) is the air volume of the chamber. This model assumes that the surface flux is in steady state such that it does not change during the chamber closure period.

3.2. Non-steady State Diffusion Model

We also explored a non-linear model based on non-steady state diffusion theory that accounts for feedbacks associated with accumulation of CO_2 in a closed chamber. The model that we use is based on the analytical solution to a partial differential equation (PDE) of soil [CO_2] dynamics that assumes the soil acts as a source of CO_2 (e.g., CO_2 is produced by microbial decomposition and root respiration). The model (PDE solution) is given in Livingston et al. [2006] as:

$$C_t = C_0 + f\tau\left(\frac{A}{V}\right)\left[\frac{2}{\sqrt{\pi}}\sqrt{t/\tau} + \exp(t/\tau)\mathrm{erfc}\left(\sqrt{t/\tau}\right) - 1\right] \tag{1.2}$$

C_0, f, A, and V are defined analogous to the corresponding terms in Eqn (1.1); τ (sec) is a "time constant" given by $\tau = (V/A)^2(\phi D_c)^{-1}$, which is a dynamic quantity that varies with soil water content via its dependence on ϕ and D_c, where ϕ (m^3 air m^{-3} soil) is the soil air-filled porosity and D_c (m^2 sec^{-1}) is the soil gas diffusion coefficient. In Eqn (1.2), erfc is the complimentary error function, which is related to the standard normal cumulative distribution function (Φ):

$$\mathrm{erfc}\left(\sqrt{t/\tau}\right) = 2\left[1 - \Phi\left(\sqrt{2t/\tau}\right)\right] \quad (1.3)$$

Eqn (1.2) assumes that horizontal transport of CO_2 within the soil is minimal, which is reasonable given the relatively short duration of our chamber sessions (30-45 min) [*Davidson et al.*, 2002] and the relatively deep insertion (8 cm) of our chambers into the soil.

Air-filled porosity, ϕ, is computed from measured volumetric soil water content (θ, m^3 m^{-3}) as:

$$\phi = 1 - \frac{BD}{PD} - \theta \quad (1.4)$$

where BD (g m^{-3}) is the soil bulk density, and PD (g m^{-3}) is the soil particle density. The diffusion coefficient, D_c, is allowed to vary in response to soil physical characteristics representative of the PHACE site [*Morgan et al.*, 2011], based on Moldrup et al. [2000]:

$$D_c = D_0\left(2\phi_{100}^3 + 0.04\phi_{100}\right)\left(\frac{\phi}{\phi_{100}}\right)^{2+3/b} \quad (1.5)$$

D_0 (m^2 sec^{-1}) is the gas diffusion coefficient in free air given the measured soil temperature (T_{soil}, K) and atmospheric pressure (P, atm), where $D_0 = D_{stp}\left(\frac{T_{soil}}{T_0}\right)^{1.75}\left(\frac{P_0}{P}\right)$, assuming D_{stp} = 0. 0000139 m^2 sec^{-1} is the gas diffusion coefficient in free air at standard temperature (T_0 = 273.2 K) and pressure (P_0 = 0.99 atm) [*Massman*, 1998]. In Eqn (1.5), ϕ_{100} (m^3 air m^{-3} soil) is the soil air-filled porosity at a soil water potential of -100 cm H_2O, and b (unitless) is a parameter describing the soil water retention curve [*Campbell and Norman*, 1998]:

$$\Psi = \Psi_e \left(\frac{\theta}{\theta_{sat}} \right)^{-b} \tag{1.6}$$

Ψ (cm H_2O) is soil water potential, Ψ_e (cm H_2O) is the air-entry potential, and θ_{sat} (m^3 m^{-3}) is the saturated soil water content. ϕ_{100} is computed by evaluating Eqn (1.4) at $\theta = \theta_{100}$, where θ_{100} is obtained by solving Eqn (1.6) for θ as a function of Ψ, and subsequently evaluating the solution at Ψ = -100 cm H_2O. Again, θ was measured in each plot (see §2.3), and we propagate uncertainty in the water retention parameters associated with Eqns (1.5) and (1.6) based on site-level results reported in Morgan et al. [2011] (for more detail, see the Supporting Information).

3.3. Non-hierarchical Statistical Model

We fit the above linear (Eqn (1.1)) and non-steady state diffusion (Eqns (1.2)–(1.6)) models to the observed chamber C versus t data via a non-hierarchical Bayesian framework, resulting in the BL and BD models, respectively. For the BD model, we also simultaneously accounted for uncertainty in the soil water retention parameters (b, Ψ_e, and θ_{sat}); see the on-line Supporting Information. The non-hierarchical framework is somewhat analogous to more traditional approaches—that employ least squares, maximum likelihood, or other optimization algorithms—that estimate C_0 and f independently for each chamber session. That is, we treat each chamber session independently such that they do not share any common parameters. Thus, for chamber session i (i = 1, 2, …, 3139) and time t (t = 0, 900, 1800 sec for 191 sessions, or t = 0, 900, 1800, 2700 sec for 2948 sessions), we assume that the observed CO_2 concentration, C^{obs} (µmol mol^{-1}), is normally distributed around the predicted (mean) concentration:

$$C_{t,i}^{obs} \sim Normal\left(C_{t,i} \frac{RT_{lab}}{1000P_{lab}}, \sigma_i^2 \right) \tag{1.7}$$

Where C (μmol m^{-3}) is based on Eqn (1.1) or Eqn (1.2) for the BL or BD model, respectively. R is the gas constant (0.08205 L atm mol^{-1} K^{-1}), and T_{lab} (293.15 K) and P_{lab} (0.74 atm) are the laboratory temperature and pressure, respectively, under which the gas samples were analyzed. C is indexed by t because it is a function of time, and by i since each chamber session is associated with its own set of parameters (i.e., f, C_0, and the observation variance, σ^2) and physical drivers (i.e., θ, T_{soil}, and P).

Within the Bayesian framework, we specified priors for the unknown parameters. To align with traditional approaches, we assumed independent, relatively non-informative (vague) priors for each session-specific parameter such that:

$$\begin{gathered} C_{0i}, f_i \sim Normal(0, B) \\ \sigma_i \sim Uniform(0, U) \end{gathered} \tag{1.8}$$

Where the values of the prior variances (B) and the upper limit of the uniform were selected to be very large (ca. 1×10^5 - 1×10^7). Since C_0 should reflect the background [CO_2] in the treatment plots, the prior for C_0 was also truncated such that values < 300 or > 4500 μmol mol^{-1} were assigned prior probabilities of zero.

The goal of this analysis is to obtain the joint posterior distribution of the model parameters, which is proportional to the likelihood times the priors. Using the bracket notation [X] and [X|Y] to indicate the marginal and conditional (on Y) probability or probability density of X [*Gelfand and Smith*, 1990], respectively, the posterior is given by:

$$\underbrace{[\mathbf{C_0}, \mathbf{f}, \boldsymbol{\sigma} \mid \mathbf{C^{obs}}]}_{posterior} \propto \underbrace{[\mathbf{C^{obs}} \mid \mathbf{C_0}, \mathbf{f}, \boldsymbol{\sigma}]}_{likelihood} \underbrace{[\mathbf{C_0}][\mathbf{f}][\boldsymbol{\sigma}]}_{priors} \tag{1.9}$$

Where $\mathbf{C^{obs}}$ is the matrix of observed chamber [CO_2], and $\mathbf{C_0}$, $\mathbf{f}$, and $\boldsymbol{\sigma}$ are the vectors of the session-level C_{0i}, f_i, and σ_i parameters, respectively. The likelihood is given by Eqn (1.7), which is linked to Eqn (1.1) for the BL model or to Eqns (1.2)-(1.6) for the BD model via the mean or predicted [CO_2] ($C_{t,i}$), and the priors are given by Eqn (1.8).

3.4. Hierarchical Statistical Model

Regardless of the fitting method (e.g., least squares, Bayesian), traditional analyses may suffer from the fact that relatively few measurements (e.g., 3-4) are made per session, and some sessions can lead to poor fits. Traditional approaches often employ an R^2 (coefficient of determination) cut-off such that sessions yielding "low" R^2 are discarded [e.g., *Hart*, 2006; *Pihlatie et al.*, 2007], and thus, estimates of the associated flux (i.e., f) are missing for these sessions. Our hierarchical specification allows the sessions to potentially borrow strength from each other—the degree to which they borrow strength depends on the magnitude of the among session variance [*Gelman et al.*, 2012]—so sessions associated with "poor" or highly variable data will be partly informed by data obtained from "good" sessions, providing estimates of the fluxes for all sessions.

We employ three assumptions to allow sessions to borrow strength from each other. First, we assume that the sessions share some common parameters. For example, we modify the likelihood in Eqn (1.7) such that the observation variance (σ^2) is assumed to vary at the level of treatment k (k = 1, 2, …, 6 levels). That is, we assume that σ^2 is similar for each session within a given treatment (thus, σ^2 is indexed by k), but that the treatments may be associated with different variances.

Second, we assume a hierarchical model for the session-specific initial or background [CO_2] (C_{0i}) and flux (f_i) parameters such that they are nested in treatments, vegetation types, and dates. That is, for treatment k (k = 1, 2, …, 6 for ct, cT, Ct, CT, cts, ctd), CO_2 treatment level k' (k' = 1 [ambient] or 2 [elevated]), vegetation type v (v = 1 [vegetated] or 2 [vegetation removed]), and date d (d = 1, 2, … 72):

$$\begin{gathered} C_{0i} \sim Normal(\hat{C}_{0k,v,d}, \hat{\sigma}^2_{k'}) \\ f_i \sim Normal(\tilde{f}_{k,v,d}, \tilde{\sigma}^2_k) \end{gathered} \quad (1.10)$$

Thus, $\hat{\sigma}^2$ describes variability in the background [CO_2] among sessions within each k by v by d combination; we assume $\hat{\sigma}^2$ varies by CO_2 treatment level given the much larger variation that is expected under experimentally applied elevated CO_2. Similarly, $\tilde{\sigma}$ describes variability in the fluxes within each combination of k, v, and d, and we allow for $\tilde{\sigma}$ to differ among the six treatment (k) levels. Since the hierarchical prior in Eqn (1.10) results in borrowing of strength and more precise estimates of C_0 and f, we did not find it necessary to constrain C_{0i} between 300 and 4500 μmol mol^{-1}, as done in the non-hierarchical models.

Third, we assigned a hierarchical prior to the $\hat{C}_{0k,v,d}$ parameters that allows for borrowing of strength among treatments, vegetation types, and dates within each CO_2 treatment level k':

$$\hat{C}_{0k,v,d} \sim Normal(\bar{C}_{0k'}, \bar{\sigma}^2_{k'}) \quad (1.11)$$

Conversely, we give independent priors to the treatment by vegetation type by date-level flux parameters ($\tilde{f}$) because these are our primary quantities of interest, and they could vary considerably across time and among treatments. Thus, we wish to avoid borrowing of strength that could lead to an underestimate of this potential variability; hence, we give independent, vague priors to each $\tilde{f}$ following Eqn (1.8):

$$\tilde{f}_{k,v,d} \sim Normal(0, B) \quad (1.12)$$

Again, B is chosen to be sufficiently large. The remaining treatment-level parameters are assigned standard, vague priors for the variances (inverse gamma distribution) and initial [CO_2]:

$$\begin{gathered}\sigma_k^2, \tilde{\sigma}_k^2, \hat{\sigma}_{k'}^2, \bar{\sigma}_{k'}^2, \sim InvGamma(a,b) \\ \bar{C}_{0k'} \sim Uniform(L,U)\end{gathered} \tag{1.13}$$

Where a and b are sufficiently small (relatively non-informative), and L and U correspond to 300 and 4500 μmol mol^{-1}, respectively.

For the HBL and HBD models, the joint posterior distribution of the model parameters is:

$$\begin{gathered}\underbrace{[\mathbf{C_0}, \hat{\mathbf{C}}_\mathbf{0}, \bar{\mathbf{C}}_\mathbf{0}, \mathbf{f}, \tilde{\mathbf{f}}, \boldsymbol{\sigma}, \hat{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\sigma}}, \bar{\boldsymbol{\sigma}} \mid \mathbf{C^{obs}}]}_{posterior} \propto \\ \underbrace{[\mathbf{C^{obs}} \mid \mathbf{C_0}, \mathbf{f}, \boldsymbol{\sigma}]}_{likelihood} \underbrace{[\mathbf{f} \mid \tilde{\mathbf{f}}, \tilde{\boldsymbol{\sigma}}][\mathbf{C_0} \mid \hat{\mathbf{C}}_\mathbf{0}, \hat{\boldsymbol{\sigma}}][\hat{\mathbf{C}}_\mathbf{0} \mid \bar{\mathbf{C}}_\mathbf{0}, \bar{\boldsymbol{\sigma}}]}_{hierarchical\ priors} \underbrace{[\bar{\mathbf{C}}_\mathbf{0}][\tilde{\mathbf{f}}][\boldsymbol{\sigma}][\hat{\boldsymbol{\sigma}}][\tilde{\boldsymbol{\sigma}}][\bar{\boldsymbol{\sigma}}]}_{priors}\end{gathered} \tag{1.14}$$

$\mathbf{C^{obs}}$, $\mathbf{f}$, and $\mathbf{C_0}$ are as described following Eqn (1.9); here, $\hat{\mathbf{C}}_\mathbf{0}$ and $\tilde{\mathbf{f}}$ are arrays of the treatment by vegetation type by date-level initial [CO_2] and CO_2 fluxes, respectively; $\bar{\mathbf{C}}_\mathbf{0}$, $\boldsymbol{\sigma}$, $\hat{\boldsymbol{\sigma}}$, $\tilde{\boldsymbol{\sigma}}$, and $\bar{\boldsymbol{\sigma}}$ are vectors of the treatment-level initial [CO_2] and the standard deviations. The likelihood is given by Eqn (1.7) with σ_i^2 replaced with σ_k^2, the hierarchical priors are given by Eqns (1.10) and (1.11), and the priors are given by Eqns (1.12) and (1.13).

3.5. Treatment Effects

Traditional approaches to estimating the surface soil CO_2 flux obtain point estimates of f then treat these as data in subsequent analysis. This approach, however, ignores the uncertainty in the f estimates. The Bayesian approach, whether hierarchical or not, can be easily extended to account for uncertainty in the f estimates, thus facilitating a more appropriate approach to subsequent analysis of f. We demonstrate this in a simple analysis that calculates all possible pairwise treatment contrasts to obtain posterior estimates of each contrast, which can be evaluated to make inferences about treatment effects. An approach to comparing f among treatments is to first compute the average f value across all plots (p_k) and dates (d) associated with global change treatment k and vegetation type v:

$$\bar{f}_{k,v} = \frac{1}{D} \sum_{d \in [2009,2011]} \left(\frac{1}{6} \sum_{p_k=1}^{6} f_{i(k,v,d)} \right) \tag{1.15}$$

Where $i(k,v,d)$ denotes the chamber session i associated with each k, v, and d. For illustrative purposes, we only consider dates between 2009 and 2011 (thus, the number of days is $D = 41$), which corresponds to the years for which the vegetated and non-vegetated plots were always measured on the same dates.

Next, we compute all possible pairwise treatment contrasts (Δ), comparing treatment level k versus k' within each vegetation type:

$$\Delta_{k,k',v} = \bar{f}_{k,v} - \bar{f}_{k',v} \tag{1.16}$$

for $k = 1, 2, \ldots, 5$ and $k' = k + 1, \ldots, 6$, resulting in 21 pairwise comparisons (15 for the vegetated plots [6×5/2] and 6 for the non-vegetated plots [4×3/2]; treatments 5 and 6 were not applied to non-vegetated plots). The treatment contrasts (Δ's) are treated as derived quantities in the Bayesian models, and posterior distributions for each Δ are obtained. One could follow the same procedure to compute contrasts between the vegetation types within each global change treatment level. Note that an advantage of a *hierarchical* Bayesian approach is that one generally does not need to correct for family-wise errors rates associated with typical multiple comparison tests [*Gelman et al.*, 2012; *Li and Shang*, 2013].

3.6. Model Comparisons

For each of the four models, we evaluated model fit by comparing the observed concentration data (C^{obs}) versus "predicted" (or "replicated") data (C^{pred}) [*Gelman et al.*, 2004] that would be generated under the same sampling distributions (e.g., Eqn (1.7) with σ_i^2 [BL and BD] or σ_k^2 [HBL and HBD]) given the predicted concentrations (C, Eqns (1.1) or (1.2)). Model fit was qualitatively evaluated by plotting C^{pred} versus C^{obs} and by computing the R^2 from a linear regression of the posterior medians of C^{pred} versus C^{obs}. We also computed model comparison indices, including the deviance information criterion, DIC [*Spiegelhalter et al.*, 2002], and posterior predictive loss, D∞ [*Gelfand and Ghosh*, 1998]. DIC is the sum of a "model fit" term (Dbar, lower values indicate better fit) and a "penalty" term representing the effective number of parameters (pD, higher values reflect a more parameter-rich model). A difference in DIC > 10 between two models provides strong support for the model with the lowest DIC [*Spiegelhalter et al.*, 2002]. Likewise, D∞ is the sum of a model fit term and a model penalty term; while a lower D∞ implies a better model, unlike DIC, there are no specific rules of thumb for differences in D∞ among candidate models [*Gelfand and Ghosh*, 1998]. However, D∞ is generally thought to be more stable or reliable than DIC, and D∞ assesses predictive performance, whereas DIC assesses explanatory performance [*Carlin et al.*, 2006].

3.7. Implementation

All four Bayesian models were implemented in OpenBUGS [*Lunn et al.*, 2009]. For each model (BL, HBL, BD, and HBD), we ran three parallel MCMC chains for sufficiently long to obtain an equivalent of >3000 effectively independent samples from their joint posteriors. Each parameter's marginal posterior distribution was summarized by its posterior median and 95% credible interval (CI), which is defined by the 2.5th and 97.5th quantiles. The OpenBUGS code and data are available from the Dryad Digital Repository (doi:10.5061/dryad.mb605) at http://dx.doi.org/10.5061/dryad.mb605.

4. Results

4.1. Model Comparisons and Model Fit

Although the BL, HBL, and HBD models fit the data equally well ($R^2 \geq 0.98$; Fig. 1A, C, D), the BD model produced more variable predictions and under-predicted [CO_2], yielding predictions close to 0 ppm for a subset of relatively high observed values, resulting in an inferior model fit ($R^2 = 0.87$). Both non-hierarchical (BL and BD) models led to highly uncertain predictions of [CO_2] such that the 95% CIs for the C^{pred} values were exceptionally wide compared to the HBL and HBD models (Figs. 1 and 2A).

The DIC and D∞ model comparison indices also provide strong support for the hierarchical models (HBL and HBD), with slightly greater support for the HBD model. The DIC values for the BL and BD models were about 3.5-9 times higher than the DICs of the HBL and HBD models, and the D∞ values were 2-3 orders of magnitude higher (Table 1). Moreover, the HBL and HBD models resulted in notably fewer effective parameters (lower pD) and thus a more parsimonious model, owing to the borrowing of strength across the dataset.

4.2. Posterior Estimates of Soil CO_2 Flux

The main goal of implementing the four models described herein was to obtain estimates of the soil CO_2 flux rate (f) associated with each chamber session. The two linear models (BL and HBL) produced similar point estimates (posterior medians) of the f values (Fig. 3A; $r = 0.97$), whereas the BD model overestimated the f values compared to its hierarchical counterpart (HBD) (Fig. 3B; $r = 0.989$, but all points fall under the 1:1 line). While the f estimates from the HBL and HBD models were highly correlated ($r = 0.995$), the HBL model underestimated the f values by ~33% compared to the HBD model (Fig. 3C). As found for the replicated data, both non-hierarchical models also produced highly uncertain estimates of f (wide 95% CIs) compared to the hierarchical models (Figs. 2B, 3, and S1).

An advantage of the hierarchical models is that they produce estimates of soil CO_2 flux rates at the level of treatment (k), vegetation type (v), and date (d), denoted by $\tilde{f}$ in Eqn (1.10), that account for variation among plots within each treatment (as captured by the treatment-specific variance term, $\tilde{\sigma}^2$, in Eqn (1.10)). Figure 4 provides example time-series of the predicted $\tilde{f}$ values obtained from the HBD model, for three different treatment combinations, showing that the soil CO_2 flux rates were fairly similar between the ambient (ct) and elevated CO_2 and warming (CT) treatments, but removal of vegetation (ct -veg) greatly reduced the flux rates in 2009-2011 (Fig. 4).

4.3. Posterior Estimates of Other Quantities

The HBL and HBD models generally produced more precise and realistic estimate of the initial (or background) [CO_2] (C_0, Eqns (1.1), (1.2), and (1.8)) compared to the two non-hierarchical models (BL and BD) (see Fig. S2). Unlike the non-hierarchical models, the hierarchical models provided direct estimates of the overall initial [CO_2] by CO_2 treatment (i.e., $\bar{C}_0$ in Eqn (1.11)). The HBL and HBD models estimated $\bar{C}_0$ to be 488.7 [483.4, 494.2] and 467.1 [463.2, 471.0] for the ambient CO_2 treatment, and 802.5 [783.8, 820.7] and 782.2 [765.0, 799.2] for the elevated CO_2 treatments, respectively.

The HBL and HBD models also quantified four potentially important variance terms, as summarized in Table S1. For example, both models indicate that the variation in the initial [CO_2] (C_{0i}) among sessions (i) within treatments (k), vegetation types (v), and dates (d) was three orders of magnitude higher in the elevated CO_2 plots compared to the ambient CO_2 plots, and ambient showed remarkably little variation in C_{0i} (e.g., posterior medians for $\hat{\sigma}_{k'}$ were < 1 μmol mol^{-1} for the ambient plots versus ca. 200 μmol mol^{-1} for the elevated CO_2 plots; Table S1). The higher spatial and temporal variation in the elevated CO_2 plots is expected given the technology used to supply CO_2 and the effect of environmental conditions (especially wind) on the spatial and temporal variability of the CO_2 concentration within an elevated CO_2 plot [*Bunce*, 2011; *Miglietta et al.*, 2001]. However, the variation in initial [CO_2] among levels of t, v, and d ($\hat{C}_{0k,v,d}$), effectively "averaging" across sessions and plots, was comparable between elevated and ambient CO_2 treatments (i.e., posterior medians for $\bar{\sigma}_{k'}$ were only ~3 times higher in the elevated plots; Table S1). Both models also indicate that variation in the CO_2 fluxes (f_i) among sessions within each k, v, and d was lowest in the ambient (control) treatment and highest for the irrigated treatments (Table S1).

4.4. Treatment Contrasts

Although this study does not focus on quantifying the effects of the different global change treatments on soil CO_2 flux (f), we demonstrate how the Bayesian approach to estimating f can be easily extended to quantify treatment effects. If uncertainty in f is rigorously accounted for, as done in the Bayesian approach, the BL model suggests that f only differed among global change treatments (within a given vegetation type) for three of the 21 comparisons (i.e., 95% CI for Δ, Eqn (1.16), did not contain zero). Conversely, the other three models (HBL, BD, and HBD) found many differences among the treatments, yielding 17-18 Δs that were different from zero. The lack of treatment differences associated with the BL model may be attributed to the highly uncertain estimates of f (wide 95% CIs for f; e.g., Fig. 3, and hence, wide 95% CI's for Δ). However, despite the wide CIs for f generated by the BD model (Fig. 3), the uncertainty in the difference among pairs of *f* values was remarkably low (narrow CIs for Δs, Fig. 5A). As one might expect, precise estimates of *f* produced by the HBL and HBD models led to tight estimates for the Δs (Fig. 5). In general, however, the direction (positive or negative) and magnitude (posterior median) of the Δs was comparable across models (very few points fall in the gray areas in Fig. 5).

5. Discussion and Conclusions

5.1. Linear versus Non-steady State Diffusion Model

Just focusing on the hierarchical models (HBL and HBD) and point estimates (here, posterior medians), the linear (HBL) model tends to underestimate *f* by ~33% (multiplicative bias) and overestimate C_0 by ~40 ppm (additive offset) relative to the HBD model. This difference is to be expected if a linear model is fit to concentration (*C*) versus time (*t*) data obtained from fairly small, static chambers that may be subject to concentration feedbacks [*Livingston et al.*, 2006; *Pedersen et al.*, 2001]. Such feedbacks would lead to an observed non-linear, decelerating relationship between observed *C* versus *t*, and a linear model would necessarily have a flatter slope compared to the initial slope near $t = 0$, which represents the surface flux (*f*) of interest. Thus, as others have also suggested [*Venterea et al.*, 2009], the linear model is not appropriate in such situations, and a non-linear model that captures the decelerating relationship is more appropriate. In particular, it would seem most appropriate to use a model based on the physics underlying the concentration feedback effects. Thus, the non-steady state diffusion model [*Livingston et al.*, 2005; 2006] would be the preferred model. This non-steady state diffusion model is easy to implement within the hierarchical Bayesian approach, and the flexibility of the coding environment (e.g., OpenBUGS, JAGS) further facilities the application of such a model (see on-line Supplemental Material). However, the HBD model can take 10 times longer to implement in software such as OpenBUGS, such that the HBL model may be preferred in situations where concentration feedbacks are minimal.

5.2. Non-hierarchical versus Hierarchical Statistical Model

An important contribution of this study is the finding that a hierarchical statistical modeling approach may be preferred over a more standard, non-hierarchical approach for estimating fluxes from non-steady state chambers that yield a limited number of observations per session. The hierarchical approach yielded much more precise estimates of all quantities of interest, such as session-level fluxes (*f*), higher-level fluxes (e.g., $\tilde{f}$), initial (background) [CO_2] (C_0), and pairwise treatment contrasts (Δ). The reason for these more precise estimates (i.e., narrower CIs) is that the hierarchical approach results in borrowing of strength (or partial pooling) [*Gelman and Hill*, 2007; *Gelman et al.*, 2012; *Ogle et al.*, 2014] such that problematic ("bad") chamber sessions (ones with low individual R^2 values) are informed by "good" chamber sessions (e.g., Fig. 6A-E). Thus, not only did the HBL and HBD models provide more precise estimates, they also yielded more biologically realistic estimates, especially for "bad" chamber sessions. Thus, the hierarchical models are not wasteful. That is, there is no need to discard "bad" session data as the borrowing of strength attribute generally ensures that the session-level *f* estimates for these sessions are reasonable, provided that there are more "good" than "bad" sessions. Additionally, in situations where all sessions produced the same amount (e.g., 4 time points) of "good" data, there is comparatively less borrowing of strength and the predicted chamber [CO_2] values align with the observed [CO_2] values for each replicate session (e.g., Fig. 6F-J), but the hierarchical structure still produces much more precise estimates than the non-hierarchical approach.

The borrowing of strength attribute associated with the hierarchical approach also results in fewer effective parameters (i.e., decreased model complexity). This essentially overcomes the problem of a potentially over-parameterized statistical model. For example, in the non-hierarchical models, three parameters (f, C_0, σ) are being estimated for each chamber session, yet there may only be 3-4 observations of C versus t per session. Thus, there is essentially 0.75-1 parameters being informed by each data point (or 1-1.33 data points per parameter), resulting in a highly over-parameterized model. In the hierarchical models, the effective number of parameters is much less such that each parameter is effectively informed by ca. 3.5-9 times as much information compared to the two non-hierarchical models (Table 1), thus increasing the information content of the C versus t data.

5.3. Post-analysis of Flux Estimates

In this study, we present a simplified example involving pairwise treatment contrasts, with the idea that these contrasts can lend insight into potential factors (i.e., treatment effects) contributing to variation in the estimated fluxes (f's). In doing so, we propagated uncertainty in the f's to the derived Δ's, allowing us to obtain posterior distributions for the Δ's. More detailed "post-analyses" of f can also be implemented to provide greater insight into the factors governing f. As an alternative to the approach described herein for evaluating Δ, one could account for uncertainty in f in the post-analyses following a general model such as:

$$\begin{gathered} E(f_i \mid Data) \sim Normal(\mu_i, \sigma_i^2) \\ \sigma_i^2 = Var(f_i \mid Data) + \sigma_{resid}^2 \\ \mu_i = M(\boldsymbol{\beta}, \mathbf{X}) \end{gathered} \qquad (1.17)$$

$E(f_i|Data)$ is the posterior mean (or expected value) of each f value (e.g., for each chamber session), conditional on the chamber data (i.e., $Data$ = C observations). In this generic example, we assume that these point estimates, $E(f_i|Data)$, are normally distributed with mean μ_i and variance σ_i^2, but other, potentially more appropriate, distributions could be employed.

One would account for uncertainty in f when specifying the variance model, such that σ_i^2 is decomposed into two terms: $Var(f_i|Data)$ is the estimated posterior variance of each f_i, and σ^2_{resid} describes the "typical" (unknown) residual variance. (A traditional approach would assume $Var(f_i|Data) = 0$, and estimate a common, residual variance.) $E(f_i|Data)$ and $Var(f_i|Data)$ are outputs generated from the HBD (or HBL) model described herein, and are thus treated as known ("data") in the post-analysis. Flexibility in modeling the factors governing f is accommodated by the model for μ_i, $M(\boldsymbol{\beta},\mathbf{X})$, which can take on any form appropriate to the particular analysis. For example, $M(\boldsymbol{\beta},\mathbf{X})$ could represent a linear or non-linear "regression" involving a set of continuous and/or categorical covariates, $\mathbf{X}$ (e.g., soil water content, soil temperature, season, treatment level, etc.), with regression coefficients (or parameters), $\boldsymbol{\beta}$. In this post-analysis, one would obtain estimates and posterior distributions of $\boldsymbol{\beta}$ and σ^2_{resid}. The posterior results for $\boldsymbol{\beta}$ incorporate the uncertainty in the f values and are used to make inferences about the factors affecting the surface fluxes.

5.4. Future Directions

We demonstrate a hierarchical, non-steady state diffusion modeling approach to estimating soil surface CO_2 efflux (e.g., f) based on C versus t data collected from non-steady state soil chambers. Our original intention was to demonstrate this approach for estimating surfaces fluxes for multiple trace gases (e.g., N_2O, CH_4, and CO_2). However, application of the approach to N_2O and CH_4 fluxes is more challenging because the soil can act as both a source and a sink for N_2O and CH_4. The non-steady state diffusion model that we adapted from Livingston et al. [2005; 2006] is only applicable to situations where the soil acts as a source. We are not aware of a comparable solution for situations where the soil acts as both a sink and/or a source. Sahoo and

Mayya [2010] offer a potential solution by solving a two-dimensional non-steady state diffusion model, but the solution is quite complicated and cannot be easily implemented in existing software packages such as OpenBUGS or JAGS. However, one could use a simpler (e.g., exponential) equation [*Hutchinson and Mosier*, 1981; *Sahoo and Mayya*, 2010] that approximates the complicated analytical solution, and our work suggests that this should be implemented in a hierarchical statistical framework.

Acknowledgements and Data

This research was supported by the US Department of Energy's Office of Science (BER), through the Terrestrial Ecosystem Science program and the National Institute for Global Environmental Change, USDA-ARS Climate Change, Soils & Emissions and GRACEnet Programs, and USDA-CSREES Soil Processes Program (2008-35107-18655). We thank Jack Morgan for project leadership, and Dan LeCain, David Smith, Erik Hardy, Valerie O'Neill, Mary Smith, Katie Tylka, Megan Nix, and Ally Eden for technical assistance. The reader can obtain the data and model code (doi:10.5061/dryad.mb605) associated with the HBD model from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.mb605).

References

Bond-Lamberty, B., and A. Thomson (2010), Temperature-associated increases in the global soil respiration record, *Nature*, *464*(7288), 579-582.

Bunce, J. A. (2011), Performance characteristics of an area distributed free air carbon dioxide enrichment (FACE) system, *Agricultural and Forest Meteorology*, *151*(8), 1152-1157.

Campbell, G. S., and J. M. Norman (1998), *An Introduction to Environmental Biophysics. Second Edition.*, 286 pp., Springer, New York.

Carlin, B. P., J. S. Clark, and A. E. Gelfand (2006), Elements of hierarchical Bayesian inference, in *Hierarchical Modeling for the Environmental Sciences: Statistical Methods and Applications*, edited by J. S. Clark and A. E. Gelfand, pp. 3-24, Oxford University Press, New York.

Davidson, E. A., K. Savage, L. V. Verchot, and R. Navarro (2002), Minimizing artifacts and biases in chamber-based measurements of soil respiration, *Agricultural and Forest Meteorology*, *113*(1-4), 21-37.

Dijkstra, F. A., D. Blumenthal, J. A. Morgan, E. Pendall, Y. Carrillo, and R. F. Follett (2010), Contrasting effects of elevated CO_2 and warming on nitrogen cycling in a semiarid grassland, *New Phytologist*, *187*(2), 426-437.

Dijkstra, F. A., J. A. Morgan, R. F. Follett, and D. R. Lecain (2013), Climate change reduces the net sink of CH_4 and N_2O in a semiarid grassland, *Global Change Biology*, *19*(6), 1816-1826.

Gelfand, A. E., and S. K. Ghosh (1998), Model choice: A minimum posterior predictive loss approach, *Biometrika*, *85*, 1-11.

Gelfand, A. E., and A. F. M. Smith (1990), Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, *85*(410), 398-409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 668 pp., Chapman and Hall/CRC Press, Boca Raton.

Gelman, A., and J. Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 623 pp., Cambridge University Press, New York.

Gelman, A., J. Hill, and M. Yajima (2012), Why we (usually) don't have to worry about mutiple comparisons, *Journal of Research on Educational Effectiveness*, *5*, 189-211.

Gelman, A., J. Hwang, and A. Vehtari (2014), Understanding predictive information criteria for Bayesian models, *Statistics and Computing*, *24*(6), 997-1016.

Hart, S. C. (2006), Potential impacts of climate change on nitrogen transformations and greenhouse gas fluxes in forests: a soil transfer study, *Global Change Biology*, *12*(6), 1032-1046.

Hutchinson, G. L., and A. R. Mosier (1981), Improved soil cover method for field measurement of nitrous oxide fluxes, *Soil Science Society of America Journal*, *45*(2), 311-316.

LeCain, D., D. Smith, J. Morgan, B. A. Kimball, E. Pendall, and F. Miglietta (2015), Microclimatic performance of a free-air warming and CO_2 enrichment experiment in windy Wyoming, USA, *PLoS ONE*, *10*(2), doi:10.1371/journal.pone.0116834.

Li, Q., and J. Shang (2013), A Bayesian hierarchical model for multiple comparisons in mixed models, *Communications in Statistics-Theory and Methods*, *44*, 5071-5090.

Livingston, G. P., G. L. Hutchinson, and K. Spartalian (2005), Diffusion theory improves chamber-based measurements of trace gas emissions, *Geophysical Research Letters*, *32*(24).

Livingston, G. P., G. L. Hutchinson, and K. Spartalian (2006), Trace gas emission in chambers: A non-steady-state diffusion model, *Soil Science Society of America Journal*, *70*(5), 1459-1469.

Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best (2009), The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine*, *28*, 3049-3082.

Massman, W. J. (1998), A review of the molecular diffusivities of H_2O, CO_2, CH_4, CO, O_3, SO_2, NH_3, N_2O, NO, and NO_2 in air, O_2 and N_2 near STP, *Atmospheric Environment*, *32*(6), 1111-1127.

Miglietta, F., M. R. Hoosbeek, J. Foot, F. Gigon, A. Hassinen, M. Heijmans, A. Peressotti, T. Saarinen, N. van Breemen, and B. Wallen (2001), Spatial and temporal performance of the MiniFACE (Free Air CO_2 Enrichment) system on bog ecosystems in northern and central Europe, *Environmental Monitoring and Assessment*, *66*(2), 107-127.

Moldrup, P., T. Olesen, P. Schjønning, T. Yamaguchi, and D. E. Rolston (2000), Predicting the gas diffusion coefficient in undisturbed soil from soil water characteristics, *Soil Science Society of America Journal*, *64*(1), 94-100.

Morgan, J. A., D. R. LeCain, E. Pendall, D. M. Blumenthal, B. A. Kimball, Y. Carrillo, D. G. Williams, J. Heisler-White, F. A. Dijkstra, and M. West (2011), C4 grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland, *Nature*, *476*(7359), 202-205.

Ogle, K., S. Pathikonda, K. Sartor, J. W. Lichstein, J. L. D. Osnas, and S. W. Pacala (2014), A model-based meta-analysis for estimating species-specific wood density and identifying potential sources of variation, *Journal of Ecology*, *102*(1), 194-208.

Parkin, T. B., and R. T. Venterea (2010), Chapter 3: Chamber-based trace gas flux measurements, in *Sampling Protocols: USDA-ARS GRACEnet project protocols.*, edited by R. F. Follett, pp. 3-1:3-39, USDA-ARS, Fort Collins, CO.

Pedersen, A. R., S. O. Petersen, and F. P. Vinther (2001), Stochastic diffusion model for estimating trace gas emissions with static chambers, *Soil Science Society of America Journal*, *65*(1), 49-58.

Pihlatie, M., J. Pumpanen, J. Rinne, H. Ilvesniemi, A. Simojoki, P. Hari, and T. Vesala (2007), Gas concentration driven fluxes of nitrous oxide and carbon dioxide in boreal forest soil, *Tellus Series B-Chemical and Physical Meteorology*, *59*(3), 458-469.

Raich, J. W., and W. H. Schlesinger (1992), The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate, *Tellus Series B-Chemical and Physical Meteorology*, *44*(2), 81-99.

Ryan, E. M., K. Ogle, T. J. Zelikova, D. R. LeCain, D. G. Williams, J. A. Morgan, and E. Pendall (2015), Antecedent moisture and temperature conditions modulate the response of ecosystem respiration to elevated CO_2 and warming, *Global Change Biology*, *21*(7), 2588-2602.

Sahoo, B. K., and Y. S. Mayya (2010), Two dimensional diffusion theory of trace gas emission into soil chambers for flux measurements, *Agricultural and Forest Meteorology*, *150*(9), 1211-1224.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *64*, 583-616.

Venterea, R. T., K. A. Spokas, and J. M. Baker (2009), Accuracy and precision analysis of chamber-based nitrous oxide gas flux estimates, *Soil Science Society of America Journal*, *73*(4), 1087-1093.

Wagner, S. W., D. C. Reicosky, and R. S. Alessi (1997), Regression models for calculating gas fluxes measured with a closed chamber, *Agronomy Journal*, *89*(2), 279-284.

Zelikova, T. J., D. G. Williams, R. Hoenigman, D. M. Blumenthal, J. A. Morgan, and E. Pendall (2015), Seasonality of soil moisture mediates responses of ecosystem phenology to elevated CO_2 and warming in a semi-arid grassland, *Journal of Ecology*, *103*(5), 1119-1130.

Table 1. Summary of model fit and comparison indices. The coefficient of determination (R^2) was obtained from a least-squares regression of the predicted (posterior median of replicated data) versus observed chamber [CO_2] data. Differences in the deviance information criterion (DIC) and posterior predictive loss (D_∞) were computed for the BL, HBL, and BD model relative to (minus) the HBD model (i.e., the HBD model had the lowest DIC and D_∞). The relative, effective number of parameters (pD) was computed for the BL, HBL and BD models as their pD values divided by the pD value for the HBD model (the HBD model had the lowest pD).

Model*	R^2	Difference in DIC	Relative pD**	Difference in D_∞
BL model	0.98	7.8×10^4	8.89	8.8×10^{10}
HBL model	0.98	4.1×10^3	3.52	7.9×10^7
BD model	0.87	6.4×10^4	1.02	1.0×10^{11}
HBD model	0.99	0	1	0

*BL = non-hierarchical Bayesian linear model; HBL = hierarchical Bayesian linear model; BD = non-hierarchical Bayesian non-steady state diffusion model; HBD = hierarchical Bayesian non-steady state diffusion model.

**We used the alternative formulation that computes pD from the posterior variance of the log likelihood [*Gelman et al.*, 2014].

Figure 1

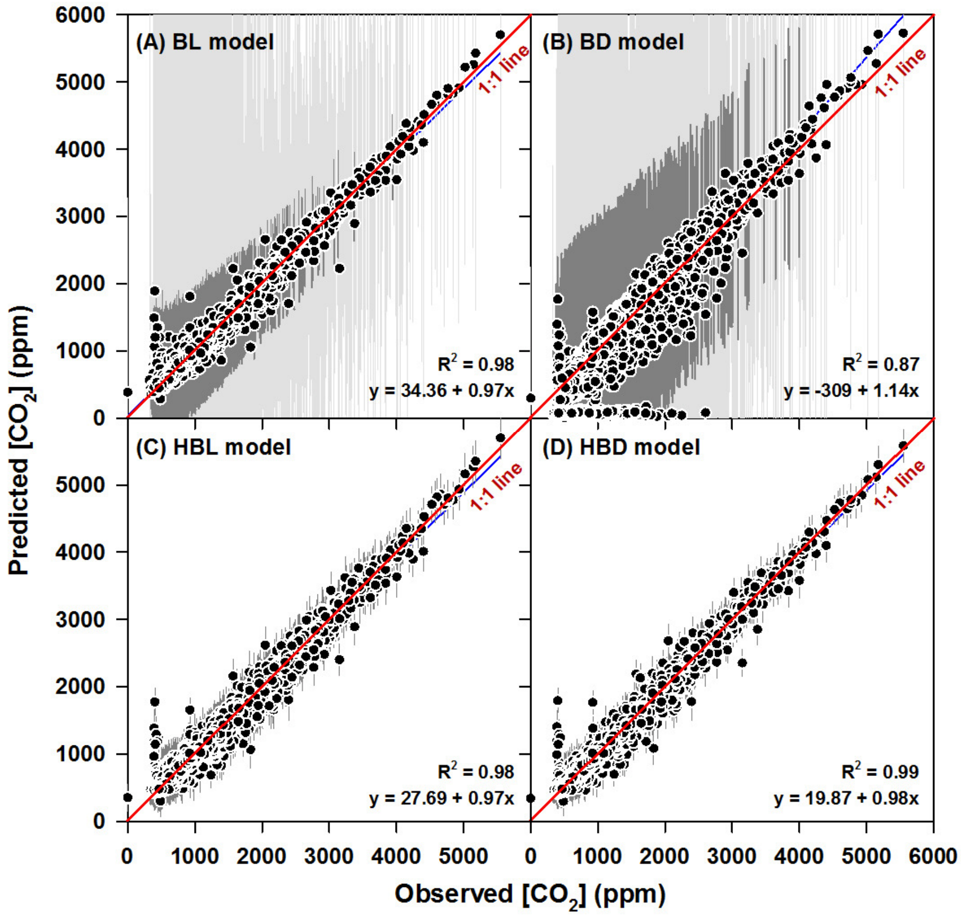

Figure 1. Observed versus predicted chamber [CO_2] for the **(A)** non-hierarchical Bayesian linear (BL) model, **(B)** non-hierarchical Bayesian, non-steady state diffusion (BD) model, **(C)** hierarchical Bayesian linear (HBL) model, and **(D)** hierarchical Bayesian, non-steady state diffusion (HBD) model. The best fit line is indicate by the thin blue diagonal line; the 1:1 line is indicated by the thick red diagonal line. Each point represents an individual observation (N = 12,240). The predicted [CO_2] values are the posterior medians (symbols) and 95% credible intervals (CIs, gray error bars) for each replicated data point. For the non-hierarchical models (BL and BD), the narrowest 50% of the CIs are indicated by dark gray, and the widest 50% are indicated by light gray.

Figure 2

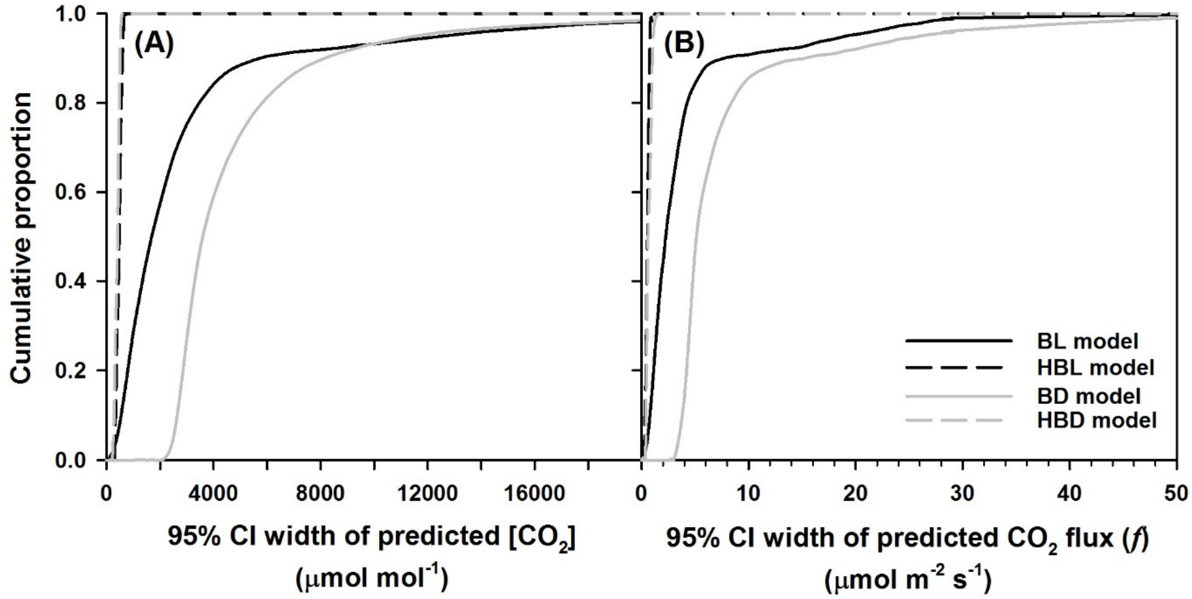

Figure 2. Cumulative distribution of the 95% CI widths for each **(A)** observation level replicated chamber [CO_2] data point (N = 12,240), and **(B)** session-level estimated soil surface CO_2 flux (N = 3139). The CI widths are computed at the 97.5th percentile minus the 2.5th percentile based on the corresponding posterior distributions. See Fig. 1 for a description of the models (BL, HBL, BD, and HBD).

Figure 3

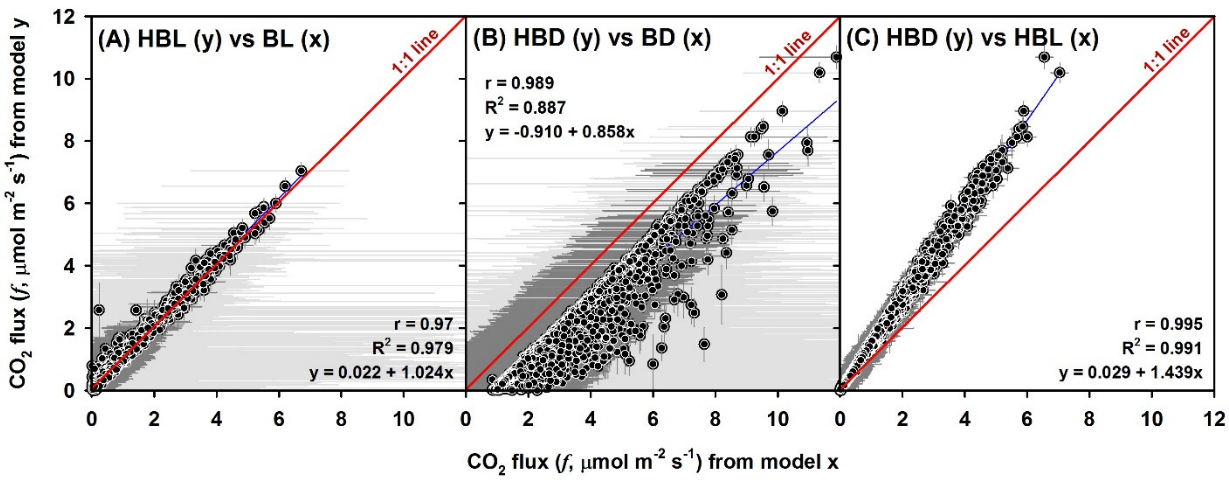

Figure 3. Comparison of the predicted session-level, surface soil CO_2 fluxes (*f*) obtained from the four models described in Figure 1 (BL, HBL, BD, and HBD). The points depict the posterior medians for each model, and the horizontal and vertical gray error bars denote the 95% CIs for the y and x models, respectively. The thin blue lines indicate the best fit line; the thick diagonal red line denotes the 1:1 line.

Figure 4

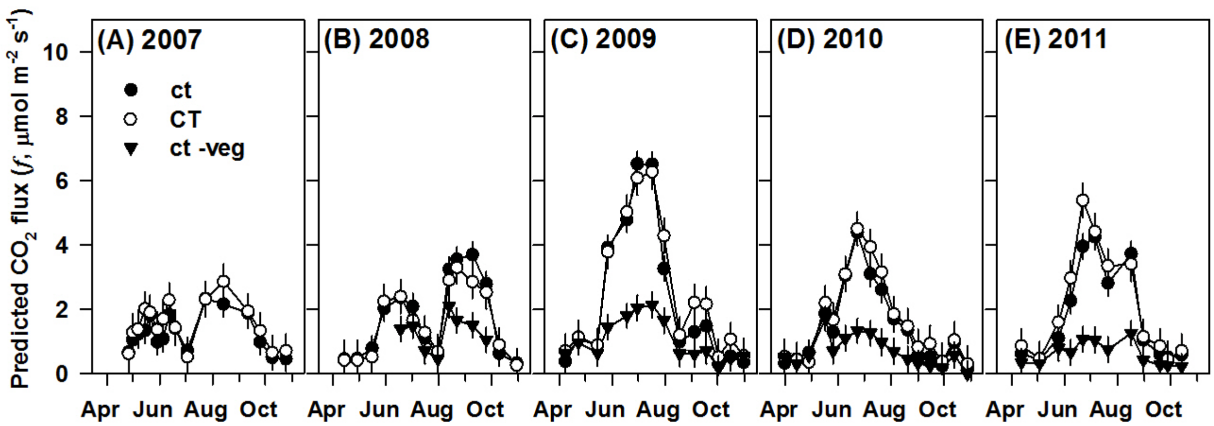

Figure 4. Predicted (posterior medians and 95% CIs) treatment-level surface soil CO_2 fluxes ($\tilde{f}$ in Eqns (1.10) and (1.12)) for a subset of treatments, for each of the five growing seasons for which chamber data were collected. The treatments shown are: ambient CO_2 and temperature (ct), elevated CO_2 and warming (CT), and ambient CO_2 and temperature with vegetation removed (ct-veg). Predictions were generated by the hierarchical Bayesian, non-steady state diffusion (HBD) model.

Figure 5

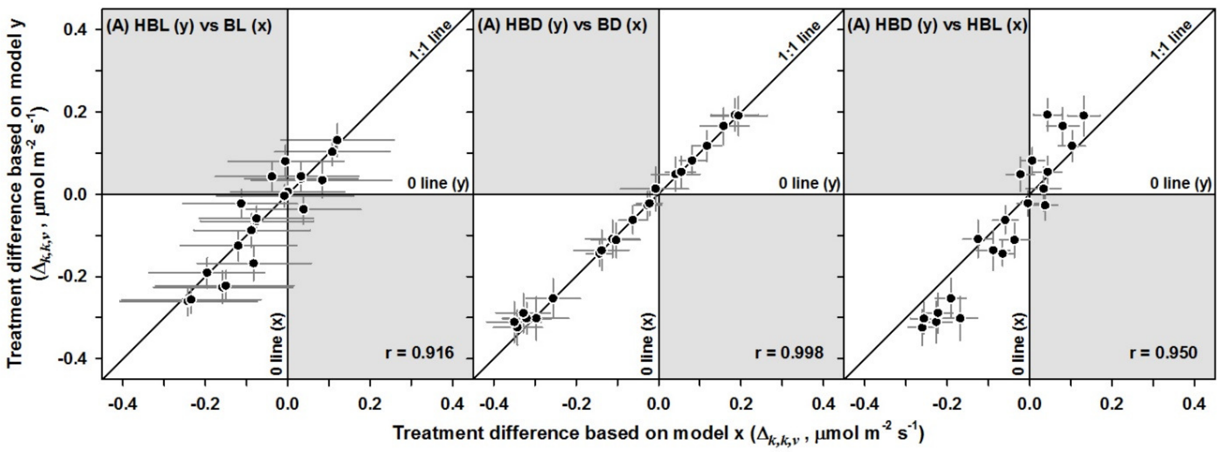

Figure 5. Comparison of the posterior estimates (medians) for the pairwise treatment contrasts (Δ, see Eqn (1.16)) between the four models described in Figure 1 (BL, HBL, BD, and HBD). The quadrats shaded in gray indicate conflicting results generated by the two models being compared (e.g., model x predicts *f* is higher for treatment *k* relative to *k*', whereas model y predicts the opposite). The white (unshaded) quadrats indicate general agreement among the two models, and points that fall along the diagonal 1:1 line indicate perfect agreement between the models, with respect to the posterior median. The BL model only yielded three Δ values that were significantly different from zero (i.e., 95% credible intervals [CIs] for a particular Δ did not contain zero), whereas the HBL, BD, and HBD models yielded 17, 17, and 18 significant Δ values, respectively.

Figure 6

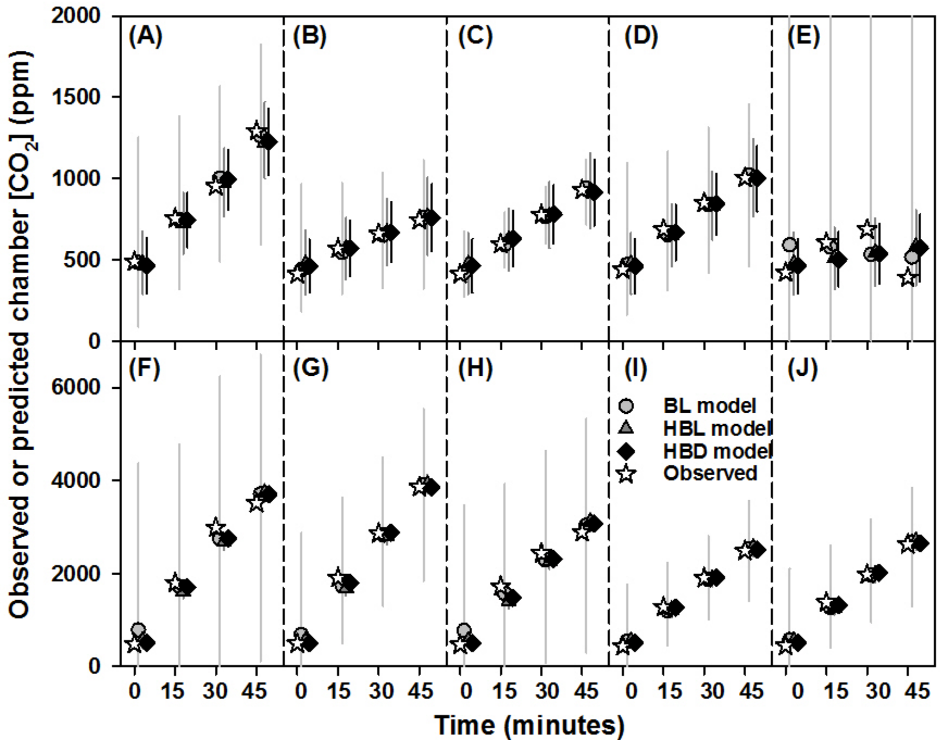

Figure 6. Example chamber sessions for **(A-E)** April 25, 2011, for the control (ct) treatment (ambient CO_2 and temperature), and **(F-J)** June 18, 2009, for the ambient CO_2 and warming (cT) treatment. Observed and predicted (posterior medians and 95% CIs) for chamber [CO_2] values are shown for each of the five replicate plots for each date, based on the BL, HBL, and HBD models (see Fig. 1 for a description of the models); results for the BD model are not shown for clarity of presentation and given its poor fit (Fig. 1B). These results demonstrate the utility of the hierarchical approach for yielding more realistic estimates of the soil surface flux (*f*) for chamber sessions associated with poor data **(E)**; for this session, the BL model predicted a negative flux, while the HBL and HBD models predicted positive fluxes that are consistent with the other sessions on that day. On dates the yielded “good” sessions for all five replicates (e.g., **F-J**), the BL, HBL, and HBD models produced similar predictions, but BL and HBL tend to slightly

overestimate the initial [CO_2]. Symbols and corresponding CIs are systematically jittered to increase visibility; some CIs are very narrow and are hidden behind their corresponding symbol.

AGU PUBLICATIONS

Journal of Geophysical Research -- Biogeosciences

Supporting Information for

Quantifying and reducing uncertainties in estimated soil CO_2 fluxes with hierarchical data-model integration

Kiona Ogle[1], Edmund Ryan[2], Feike Dijkstra[3], Elise Pendall[4]

[1]School of Informatics, Computing & Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA; [2]Lancaster Environment Center, Lancaster University, Lancaster, UK; [3]Centre for Carbon, Water & Food, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia; [4]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, Australia

Contents of this file

Text S1
Table S1
Figures S1 and S2

Introduction

This supporting information provides a description of how the soil water retention parameters were estimated and incorporated into the non-steady state diffusion model (Text S1).

Supporting figures are also included that illustrate (1) uncertainties associated with the estimated soil surface CO_2 fluxes obtained under four different modeling approaches (Fig. S1), and (2) estimates of background or initial chamber [CO_2] obtained from the four different modeling approaches, for representative plots, dates, and global change treatments (Fig. S2).

Text S1: Soil Water Retention Parameters

The soil water retention parameters (θ_{100} and b) relevant to the model for D_c (CO_2 diffusion coefficient) were estimated by fitting the log-transformed version of Eqn 1.6 to data on soil water content (θ) and soil water potential (Ψ). Direct measurements of θ were made at the field site, while Ψ was estimated using soil texture data measured at the field site and pedotransfer functions in the Rosetta software (version 1.2) [see supplemental materials in, *Morgan et al.*, 2011]. We fit Eqn 1.6 to the θ and Ψ data within a simple, non-hierarchical Bayesian framework and assigned uniform, U(0, 100), priors to each of b and the intercept (θ_{100} is a deterministic function of b and the intercept), and we estimated the values of these parameters at the site level. The Bayesian analysis was implemented in OpenBUGS [*Lunn et al.*, 2009], which adopts a Markov chain Monte Carlo (MCMC) approach to approximate the joint posterior distribution of the parameters. This produced 3000 independent samples of a and θ_{100} from the posterior. The posterior means and variances from this analysis were used to specify informative priors for b and θ_{100} within the non-steady state diffusion models (i.e., for both the BD and HBD models). In particular, we assumed $\theta_{100} \sim Normal(0.374,0.000201)I(0.348,0.404)$ and $b \sim Normal(4.55, 0.2058)I(4.23,4.86)$, where $I(A,B)$ indicates that the normal distribution was truncated at A and B such that parameter values lying outside these limits are associated with zero probability density. The values for A and B were set equal to the 2.5th and 97.5th percentiles obtained from the Bayesian analysis of the soil water retention data.

References

Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best (2009), The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine*, *28*, 3049-3082.

Morgan, J. A., D. R. LeCain, E. Pendall, D. M. Blumenthal, B. A. Kimball, Y. Carrillo, D. G. Williams, J. Heisler-White, F. A. Dijkstra, and M. West (2011), C4 grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland, *Nature*, *476*(7359), 202-U101.

Parameter	Treatment level	HBL model 2.5th	HBL model Median	HBL model 97.5th	HBD model 2.5th	HBD model Median	HBD model 97.5th
σ_1	ct	90.7	93.8	97.0	80.2	82.8	85.7
σ_2	cT	76.1	78.6	81.4	59.1	61.1	63.2
σ_3	Ct	105.4	109.7	114.3	88.2	91.9	95.7
σ_4	CT	119.9	124.8	129.9	105.8	110.3	114.9
σ_5	ctd	103.2	107.7	112.8	82.8	86.4	90.4
σ_6	cts	116.8	121.9	127.5	95.5	99.7	104.3
$\hat{\sigma}_1$	c	0.02	0.52	3.16	0.01	0.14	0.55
$\hat{\sigma}_2$	C	188.7	199.6	211.0	193.9	204.5	215.8
$\bar{\sigma}_1$	c	34.3	39.2	44.5	21.4	25.6	30.0
$\bar{\sigma}_2$	C	84.2	102.5	121.7	77.8	95.9	114.4
$\tilde{\sigma}_1$	ct	0.28	0.30	0.32	0.39	0.41	0.44
$\tilde{\sigma}_2$	cT	0.36	0.39	0.41	0.53	0.57	0.60
$\tilde{\sigma}_3$	Ct	0.38	0.41	0.45	0.54	0.58	0.62
$\tilde{\sigma}_4$	CT	0.35	0.38	0.41	0.50	0.54	0.58
$\tilde{\sigma}_5$	ctd	0.54	0.59	0.65	0.76	0.82	0.89
$\tilde{\sigma}_6$	cts	0.67	0.73	0.79	1.02	1.11	1.21

Table S1. Posterior estimates (median and 95% credible interval limits, 2.5th and 97.5th percentiles) of the standard deviation terms associated the two hierarchical Bayesian models: HBL (linear process model) and HBD (non-steady state diffusion model). Treatment codes are: ct = ambient CO_2 and temperature, cT = ambient CO_2 and warming, Ct = elevated CO_2 and temperature, CT = elevated CO_2 and warming, cts = shallow irrigation, ctd = deep irrigation, c = ambient CO_2, and C = elevated CO_2. See text following Eqn 1.14 for a description of σ (the [CO_2] residual error variance); see Eqn 1.10 for a description of $\hat{\sigma}$ and $\tilde{\sigma}$ and Eqn 1.11 for a description of $\bar{\sigma}$; σ, $\hat{\sigma}$, and $\bar{\sigma}$ have units of μmol mol^{-1}, and $\tilde{\sigma}$ has units of μmol m^{-2} s^{-1}.

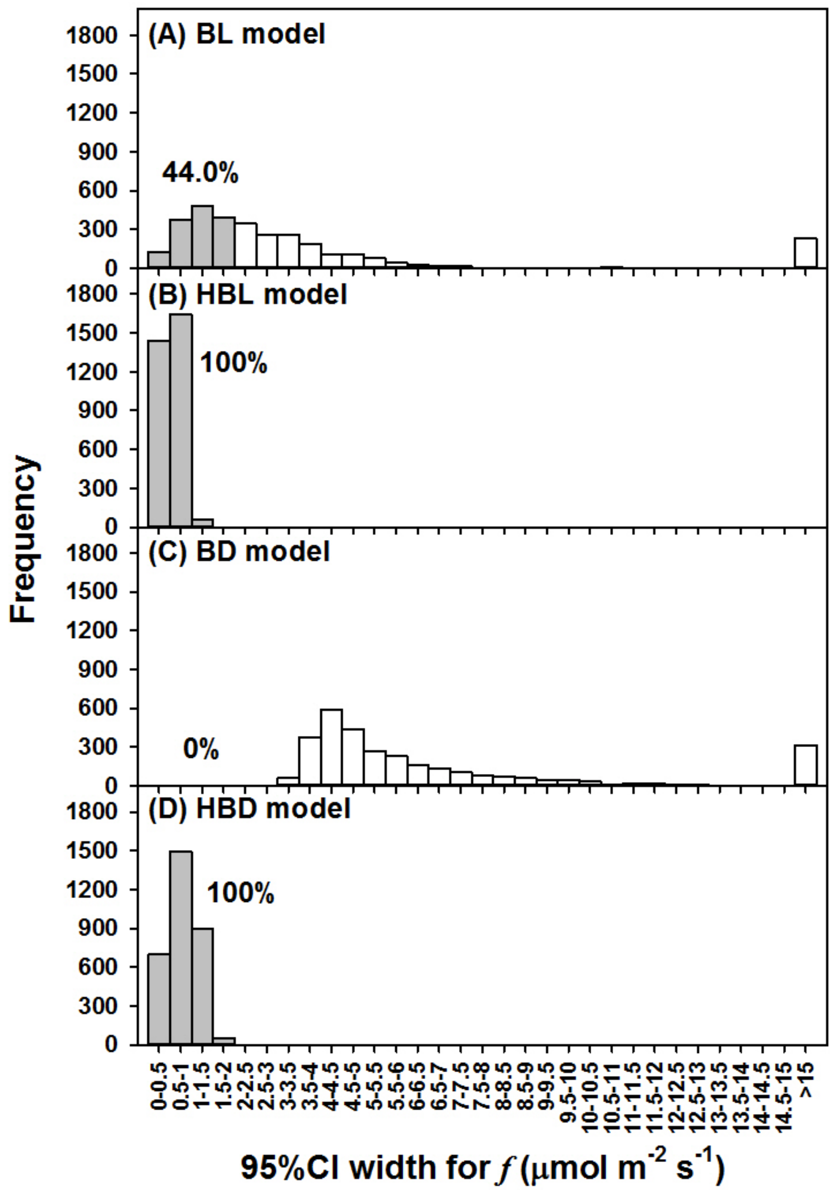

Figure S1. Histograms of the session-level 95% credible interval (CI) widths for the soil surface CO_2 flux (*f*) obtained from the (A) non-hierarchical Bayesian, linear (BL) model, (B) hierarchical Bayesian, linear (HBL) model, (C) non-hierarchical Bayesian, non-steady state diffusion (BD) model, and (D) hierarchical Bayesian, non-steady state diffusion (HBD) model. Bars shaded in gray indicate CI widths <2 μmol m^{-2} s^{-1}, and the percentages near the shaded bars indicate the percent of sessions (out of 3139) associated with CI widths <2 μmol m^{-2} s^{-1}.

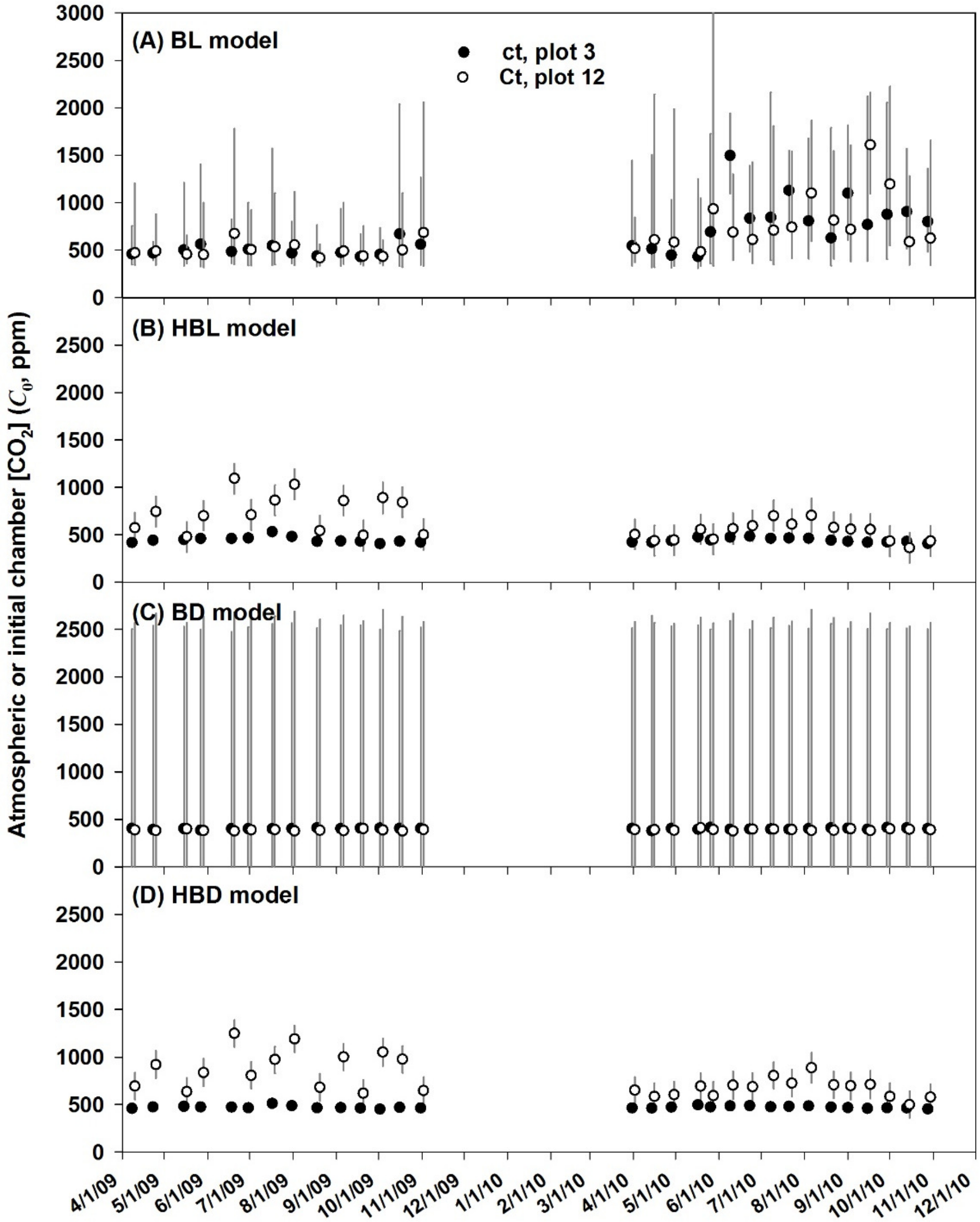

Figure S2. Examples of the predicted initial chamber [CO_2] at time $t = 0$ (i.e., C_0 in Eqn (1.2)). Representative examples are shown for one plot in each of the ct (ambient CO_2, ambient temperature) and Ct (elevated CO_2, ambient temperature) treatments, for the (A) non-hierarchical Bayesian, linear (BL) model, (B) hierarchical Bayesian, linear (HBL) model, (C) non-hierarchical Bayesian, non-steady state diffusion (BD) model, and (D) hierarchical Bayesian, non-steady state diffusion (HBD) model. The symbols denote the posterior medians, and the error bars denote the 95% credible intervals (CIs). The results for the non-hierarchical (BL and BD) models resulted in wider 95% CIs than the hierarchical (HBL and HBD) models.