Agricultural and Forest Meteorology 149 (2009) 1552–1555

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Short communication

Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes

Dean Vickers[a,*], Christoph K. Thomas[a], Jonathan G. Martin[b], Beverly Law[b]

[a] College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA
[b] College of Forestry, Oregon State University, Corvallis, OR, USA

ARTICLE INFO

Article history:
Received 3 October 2008
Received in revised form 5 February 2009
Accepted 21 March 2009

Keywords:
Carbon dioxide flux
Eddy-covariance
Flux partitioning
Self-correlation

ABSTRACT

Self-correlation between estimates of assimilation and respiration of carbon is a consequence of the flux partitioning of eddy-covariance measurements, where the assimilation is computed as the difference between the measured net carbon dioxide flux (*NEE*) and an estimate of the respiration. The estimates of assimilation and respiration suffer from self-correlation because they share a common variable (the respiration). The issue of self-correlation has been treated before, however, published studies continue to report regression relationships without accounting for the problem. The self-correlation is defined here (for example) as the correlation between variables *A* and *B*, where $A = x + y$ and $B = x$, and where *x* and *y* are random, uncorrelated variables (random permutations of the observations). In this case, any correlation found between *A* and *B* has no physical meaning and is entirely due to the self-correlation associated with the shared variable *x*. Estimates for the self-correlation are presented for a range of timescales using two different methods applied to a 6-yr dataset of eddy-covariance and soil chamber measurements from a ponderosa pine forest. Although the estimate of self-correlation is itself uncertain, it is not small compared to the observed correlation, and therefore it can reduce the strength of the relationship that can be demonstrated even though there is a strong apparent relationship in the observations and a strong causal relationship is expected based on tree physiology through coupling of photosynthesis and autotrophic respiration. We conclude that previous studies using eddy-covariance measurements and standard flux partitioning methods may have inadvertently overstated the real correlation between assimilation and respiration because they failed to account for self-correlation.

1. Introduction

Partitioning of the net ecosystem exchange (*NEE*) of carbon, the quantity measured by eddy-covariance, into gross ecosystem production (*GEP*, or assimilation) and ecosystem respiration (*ER*) is necessary to improve understanding of the causes of interannual variability and between-site variability of annual *NEE*, and for verification and improvement of process-based carbon models. Establishing relationships between *GEP* and *ER* is important for predicting the influence of climate change on future sequestration of carbon by ecosystems. For example, observational studies often report that *GEP* and *ER* are strongly correlated (e.g., Janssens et al., 2001; Hogberg et al., 2001; Reichstein et al., 2007; Stoy et al., 2008; Baldocchi, 2008). Such correlation would imply that potential changes in *GEP* associated with climate change may be partially offset by corresponding changes in *ER*, and that therefore, the *NEE* may be relatively insensitive to climate change. However, evaluation of this hypothesis using eddy-covariance must consider the problem of self-correlation. Self-correlation (Hicks, 1978; Klipp and Mahrt, 2004; Baas et al., 2006) has also been referred to as spurious correlation (Pearson, 1897; Kenney, 1982; Jackson and Somers, 1991; Brett, 2004) and as the shared variable problem in the statistics literature. It arises when one group of variables is plotted against another, and the two groups have one or more variables in common. For example, $x + y$ and x are self-correlated because they share the common variable *x*. For variables suffering from self-correlation, the coefficient of determination is not directly related to the quality of the data or to the validity of the relationship being considered.

The self-correlation must be taken into account when interpreting observations to develop relationships (e.g., Perrie and Toulany, 1990; Mahrt and Vickers, 2003; Hsu and Blanchard, 2004; Lange et al., 2004; Klipp and Mahrt, 2004; Mauritsen and Svensson, 2007). Normally a small correlation indicates no relationship between the variables being compared, however, when the variables tested share a common variable or factor, even random

* Corresponding author at: College of Oceanic and Atmospheric Sciences, COAS Admin Bldg 104, Oregon State University, Corvallis, OR 97331-5503, USA. Tel.: +1 541 737 5706.
E-mail address: vickers@coas.oregonstate.edu (D. Vickers).

0168-1923/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2009.03.009

data can produce large correlation (Kim, 1999). This self-correlation is a better reference point of no real relationship than $R^2 = 0$. Following Klipp and Mahrt (2004), we take the difference between the variance explained by the observations and the variance explained by self-correlation as a measure of the variance explained by real underlying processes. The difference is not a true variance in that it can be negative. In this study we calculate the self-correlation between *GEP* and *ER* using 6 years of eddy-covariance and chamber measurements collected at a pine forest site. The flux partitioning is performed for two approaches for estimating respiration: (1) using nocturnal eddy-covariance measurements to develop time- and temperature-dependent models of *ER* which are then extrapolated to the daytime, and (2) using automated soil chamber measurements to estimate *ER*. The self-correlation is computed using two methods: (1) a general method that assumes no special relationships between the variables, and (2) a more refined method specific to CO_2 flux partitioning. Results are presented for a range of averaging timescales.

2. Methods

2.1. Site description and measurements

The site is a semi-arid, 90-yr-old ponderosa pine forest in Central Oregon, U.S.A. (Irvine et al., 2008). The pine canopy extends from 10 to 16 m above ground, while the understory consists of scattered 1-m tall shrubs. The leaf area index (LAI) ranges from 3.1 to 3.4 during the growing season and the stand density is 325 trees ha^{-1}. Although the site is located on a relatively flat saddle region about 600 m across, it is surrounded by complex terrain.

Eddy-covariance measurements were collected using a Campbell Scientific CSAT3 sonic anemometer and an open-path LICOR-7500 gas analyzer at 33 m above ground (or about twice the canopy height). Additional measurements include vertical profiles of air temperature (HOBO thermistors) and mean CO_2 concentration using a LICOR-6262 with inlets at 1, 3, 6, 15 and 33 m above ground. The CO_2 profile system was replaced by a new system with a LICOR-820 gas analyzer with inlets at 0.3, 1, 3, 6, 10, 18 and 33 m above ground in August of 2006. An estimate of ecosystem respiration based on chamber measurements was made by combining high temporal resolution (1-h average) data from an automated soil respiration system (Irvine and Law, 2002) with estimates of foliage and live wood respiration derived from temperature response functions specific to ponderosa pine (Law et al., 1999). The soil chamber estimates include the respiration from fine woody debris. Extensive periodic manual soil respiration measurements from a LICOR-6400 with a LICOR-6000-9 soil chamber were used to scale the automated chamber measurements as described in Irvine et al. (2008). The data analyzed here were collected during 2002 through 2007.

2.2. Data processing

A brief overview of the data processing and gap-filling techniques is presented here. Raw 10/20 Hz eddy-covariance data and 30-min fluxes and variances were subjected to quality control based on a combination of tests for plausibility, stationarity and well-developed turbulence (Foken et al., 2004). For nighttime periods, gaps in the *NEE* time series were filled using a temperature response (Arrhenius-type) model with separate coefficients for different soil moisture categories, and for the daytime, gaps in the *NEE* were filled using a light response model with separate coefficients for different temperature and soil moisture classes (Ruppert et al., 2006). The daytime *NEE* estimates were partitioned into assimilation and respiration components by extrapolating the nighttime approach into daytime conditions. *NEE* was calculated as the sum of the eddy-covariance flux and the storage term. To gap-fill the respired CO_2 from the soil chamber measurements, a multivariable regression approach was used where the predictor variables include air temperature, soil temperature, net radiation, soil moisture and vapor pressure deficit.

2.3. Self-correlation: Method I

With the gap-filled estimates of *ER* and *NEE* based on measurements and/or modeling as described above, the *GEP* is calculated as a residual

$$GEP = NEE - ER \tag{1}$$

to balance the budget, and therefore, *GEP* and *ER* are self-correlated because they both contain *ER*. Such self-correlation is the same sign as the expected correlation, and if large enough can lead to false confidence in the hypothesis that variations of *GEP* and *ER* are tightly coupled.

We define the self-correlation using the following procedure after Klipp and Mahrt (2004). A uniform random number generator is used to create a random sequence of length *N* of integer values between 1 and *N*, where *N* is the length of the observed series. Random permutations of the 24-h sums of *NEE* and *ER* are created using the random sequence as index pointers into the pool of observed values. Given the random and uncorrelated permutations of *NEE* and *ER*, the *GEP* is calculated as a residual and the correlation is computed between *GEP* and *ER*. Because the random permutations no longer retain any real connections between *GEP* and *ER*, the correlation computed from such series has no physical meaning and is a measure of the self-correlation due to *GEP* and *ER* sharing a common variable (Hicks, 1978; Andreas and Hicks, 2002; Mahrt and Vickers, 2003; Klipp and Mahrt, 2004; Baas et al., 2006). This process is repeated for many realizations to construct the probability distribution of the self-correlation. Using randomized actual data to evaluate the self-correlation, rather than data synthesized by a random number generator, is preferred because the frequency distribution of the observed data is reproduced in the random permutation (Kim, 1999).

2.4. Kenney (1982)

Here we briefly discuss the work by Kenney (1982) on spurious correlation and reconcile it with our numerical approach described above. Kenney (1982) defines variables $A = x + y$ and $B = x$, where x and y are measured quantities and the relationship between A and B is of interest. In terms of such a relationship, x is a shared variable and therefore some or all of the correlation between A and B may be spurious. He then develops the expression for the correlation between A and B in terms of the variances and covariance of x and y (r_{AB}) in his Eq. (5). However, at this point in his development Kenney (1982) generates some confusion (in our opinion) by calling r_{AB} the "spurious self-correlation coefficient", while in fact, it is simply the correlation between A and B. We would prefer different wording to make it clear that some fraction of the correlation between A and B may be spurious, but not all of it. This confusion may have led to some of the past discussion on this issue (Prairie and Bird, 1989; Kenney, 1991).

In terms of our notation (Tables 1 and 2), Eq. (5) for r_{AB} in Kenney (1982) is equivalent to our R_{OBS}, which is simply the observed correlation between A and B, or in our case between *GEP* and *ER*. His Eq. (6) is our R_{SC}, or the self-correlation, which is the correlation between A and B when x and y (our *NEE* and *ER*) are random uncorrelated permutations of the observations. R_{SC} is a measure of the self-correlation because when there are no real