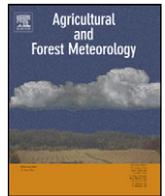




Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

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 CO₂ 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.

© 2009 Elsevier B.V. All rights reserved.

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; Mahr 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).

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

$$GEP = NEE - ER \quad (1)$$

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

Table 1

R^2 values for the observed correlation (OBS), the self-correlation (SC) and the real correlation between GEP and ER . Respiration is estimated based on nocturnal eddy-covariance data. All R^2 values are significant at the $p = 0.05$ significance level.

Timescale	N	Method I			Method II		
		R^2_{OBS}	R^2_{SC}	R^2_{REAL}	R^2_{OBS}	R^2_{SC}	R^2_{REAL}
1 day	2190	0.62	0.54	0.08	0.62	0.33	0.29
1 week	312	0.72	0.60	0.12	0.72	0.37	0.35
1 month	72	0.79	0.64	0.15	0.79	0.39	0.40
3 months	24	0.80	0.66	0.14	0.80	0.40	0.40

connections between x and y (NEE and ER) any correlation found between A and B (GEP and ER) can arise only from the shared variable problem.

Kenney (1982) showed that the magnitude of the self-correlation for this particular example depends only on the ratio of the variance of x (the shared variable) and y (the unique variable),

$$r_{SC} = \frac{1}{(1 + (S_y/S_x)^2)^{1/2}}, \quad (2)$$

when x and y are random uncorrelated permutations and S denotes the standard deviation. The self-correlation is greater when there is larger dispersion (variance) for the common term compared to the unique term. The expected value of the self-correlation from the numerical method described above agrees with Eq. (2).

2.5. Self-correlation: Method II

In this section we describe an alternate method to estimate the self-correlation specifically for CO_2 flux partitioning by taking advantage of a special characteristic of GEP . Consider replacing the 24-h sums used in Method I with sums over daytime and nighttime hours, in which case Eq. (1) becomes

$$GEP_D + GEP_N = NEE_D + NEE_N - ER_D - ER_N \quad (3)$$

where subscript D denotes a sum over daytime hours, and N denotes a sum over nighttime hours. Using the fact that GEP_N is zero (no photosynthesis in the absence of light), and that NEE_N is equivalent to ER_N (again because GEP_N is zero), reduces the above expression to

$$GEP_D = NEE_D - ER_D, \quad (4)$$

which potentially involves only daytime measurements, however, this distinction is somewhat blurred when ER_D is estimated based on temperature-dependent models tuned to the nocturnal eddy-covariance measurements.

The self-correlation between GEP and ER in Method II is computed using the same numerical approach described above, however, here the random uncorrelated permutations are generated for NEE_D , ER_D and ER_N , and the correlation is computed between $(NEE_D - ER_D)$ and $(ER_D + ER_N)$. In Method II, there are two unique variables (NEE_D and ER_N) and the single shared variable is ER_D . In Method I, there is one unique variable (NEE) and a single shared variable (ER).

Table 2

Same as Table 1 except respiration is estimated using automated chamber data.

Timescale	N	Method I			Method II		
		R^2_{OBS}	R^2_{SC}	R^2_{REAL}	R^2_{OBS}	R^2_{SC}	R^2_{REAL}
1 day	2190	0.65	0.53	0.12	0.65	0.31	0.34
1 week	312	0.75	0.60	0.15	0.75	0.35	0.40
1 month	72	0.81	0.64	0.17	0.81	0.37	0.44
3 months	24	0.82	0.64	0.18	0.82	0.38	0.44

3. Results

The self-correlation is significantly larger than zero and not small compared to the observed correlation (Table 1). This result is found for both Methods I and II, although the self-correlation is substantially reduced using Method II. Averaging over all timescales tested, the variance explained by self-correlation (R^2_{SC}) is about 80% of the observed variance explained for Method I, and about 50% for Method II. That is, about one-half of the observed variance of ER explained by GEP is due to the shared variable problem even for Method II, and as a consequence, the real variance explained (R^2_{REAL} in Table 1) is about one-half of the observed. For example, although the observed variance explained between GEP and ER reaches as high as $R^2 = 0.82$ for the approach using the chamber data for a 3-month timescale, our Method II analysis indicates that 46% of that observed variance explained is due to self-correlation.

The magnitude of the self-correlation is not strongly sensitive to the choice of using nocturnal eddy-covariance or chamber measurements to estimate the respiration, although there are subtle differences for this site (Table 2). The R^2_{REAL} values are slightly larger when using the automated chamber data because the variance of the shared variable is smaller. The observed variance explained (R^2_{OBS}) is slightly larger for the chamber approach, and the majority of this additional observed variance explained appears to be real, and not associated with self-correlation.

The Method II estimates of self-correlation are smaller than those from Method I primarily because the variance of the shared variable is considerably smaller in Method II. The variance of ER (the shared variable in Method I) is greater than the variance of ER_D (shared variable in Method II) in part because ER_D and ER_N are positively correlated ($r = 0.75$). That is, 24-h periods with large daytime respiration are also associated with large nighttime respiration.

The variance explained by the observed data (R^2_{OBS}) increases with increasing timescale for the range of timescales tested here, however, so does the variance explained due to self-correlation (Table 1). Because R^2_{OBS} increases slightly faster with increasing timescale than does R^2_{SC} , the variance explained that can be attributed to real underlying processes tends to increase with increasing timescale. R^2 values found for the annual timescale (not shown) are not statistically significant in part because of the limited sample size ($N = 6$ yr).

4. Conclusions

Current standard methods of CO_2 flux partitioning applied to eddy-covariance measurements result in levels of self-correlation between assimilation and respiration that are too large to ignore (Tables 1 and 2). The self-correlation is due to the shared variable problem. At the site studied here, the self-correlation is slightly smaller using estimates of respiration based on automated chamber data compared to models based on nocturnal eddy-covariance data, and the estimates of self-correlation are substantially reduced by partitioning into daytime and nighttime periods and noting that assimilation is zero at night. The reduction in self-correlation found by the day/night partitioning method can be explained by the smaller variance of the shared variable. With or without the day/night partitioning, and for both approaches to estimating respiration (eddy-covariance or chambers), and for all averaging timescales tested, our results indicate that the self-correlation is not small.

Although the estimates of self-correlation are uncertain, and the results found here may not apply to all flux tower sites, we conclude that the self-correlation should be accounted for when

estimating the strength of the relationship between *GEP* and *ER*. Our main conclusion does not suggest that there is no strong causal relationship between assimilation and respiration, only that when such a relationship is based on eddy-covariance measurements and standard flux partitioning methods one must consider the magnitude of the self-correlation.

Acknowledgments

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant DE-FG02-06ER64318. The two reviewers are gratefully acknowledged for their suggestions to improve the manuscript. We also thank Dennis Baldocchi and Markus Reichstein for their contributions.

References

- Andreas, E.L., Hicks, B.B., 2002. Comments on "Critical test of the validity of Monin-Obukov similarity during convective conditions". *J. Atmos. Sci.* 59, 2605–2607.
- Baas, P., Steeneveld, G.J., Van De Wiel, B.J.H., Holtslag, A.A.M., 2006. Exploring self-correlation in flux-gradient relationships for stably stratified conditions. *J. Atmos. Sci.* 63, 3045–3054.
- Baldocchi, D., 2008. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Austr. J. Bot.* 56, 1–26.
- Brett, M.T., 2004. When is correlation between non-independent variables "spurious"? *Oikos* 105, 647–656.
- Foken, T., Göckede, M., Mauder, M., Mahrt, L., Amiro, B., Munger, W., 2004. Post-field data quality control. In: Lee, X., Massman, W., Law, B. (Eds.), *Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis*. Kluwer Academic Publishers, Boston, p. 250.
- Hicks, B.B., 1978. Some limitations of dimensional analysis and power laws. *Boundary-Layer Meteorol.* 14, 567–569.
- Hogberg, P., Nordgren, A., Buchmann, N., Taylor, A., Ekblad, A., Hogberg, M., Nyberg, G., Ottosson-Lofvenius, M., Read, D., 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature* 411, 789–792.
- Hsu, S.A., Blanchard, B.W., 1911–1916. Estimating overwater turbulence intensity from routine gust-factor measurements. *J. Appl. Meteorol.* 43.
- Irvine, J., Law, B.E., 2002. Contrasting soil respiration in young and old-growth ponderosa pine forests. *Global Change Biol.* 8, 1183–1194.
- Irvine, J., Law, B.E., Martin, J.G., Vickers, D., 2008. Interannual variation in soil CO₂ efflux and the response of root respiration to climate and canopy gas exchange in mature ponderosa pine. *Global Change Biol.* 14, 1–12.
- Jackson, D.A., Somers, K.M., 1991. The spectre of spurious correlations. *Oecologia* 86, 147–151.
- Janssens, I.A., Lankreijer, H., Matteucci, G., Kowalski, A.S., Buchmann, N., Epron, D., Pilegaard, K., Kutsch, W., Longdoz, B., Grunwald, T., Montagnani, L., Dore, S., Rebmann, C., Moors, E.J., Grelle, A., Rannik, U., Morgenstern, K., Oltchev, S., Clement, R., Guomundsson, J., Minerbi, S., Berbigier, P., Ibrom, A., Moncrieff, J., Aubinet, M., Bernhofer, C., Jensen, N.O., Vesala, T., Granier, A., Schulze, E., 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.
- Kennedy, B.C., 1982. Beware of spurious self-correlation! *Water Resour. Res.* 18, 1041–1048.
- Kennedy, B.C., 1991. Comments on "Some misconceptions about the spurious correlation problem in the ecological literature by Y.T. Prairie and D. F. Bird". *Oecologia* 86, 152.
- Kim, J.H., 1999. Spurious correlation between ratios with a common divisor. *Stat. Probab. Lett.* 44, 383–386.
- Klipp, C.L., Mahrt, L., 2004. Flux-gradient relationship, self-correlation and intermittency in the stable boundary layer. *Q. J. R. Meteor. Soc.* 130, 2087–2104.
- Lange, B., Johnson, H.K., Larsen, S., Hojstrup, J., Kofoed-Hansen, H., Yelland, M.J., 2004. On detection of a wave age dependency for the sea surface roughness. *J. Phys. Oceanogr.* 34, 1441–1458.
- Law, B.E., Ryan, M.G., Anthoni, P.M., 1999. Seasonal and annual respiration of a ponderosa pine ecosystem. *Global Change Biol.* 5, 169–182.
- Mahrt, L., Vickers, D., 2003. Formulation of turbulent fluxes in the stable boundary layer. *J. Atmos. Sci.* 60, 2538–2548.
- Mauritsen, T., Svensson, G., 2007. Observations of stably stratified shear-driven atmospheric turbulence at low and high Richardson numbers. *J. Atmos. Sci.* 64, 645–655.
- Pearson, K., 1897. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. R. Soc. Lond.* 60, 489–502.
- Perrie, W., Toulany, B., 1990. Fetch relations for wind-generated waves as a function of wind-stress scaling. *J. Phys. Oceanogr.* 20, 1666–1681.
- Prairie, Y.T., Bird, D.F., 1989. Some misconceptions about the spurious correlation problem in the ecological literature. *Oecologia* 81, 285–288.
- Reichstein, M., Papale, D., Valentini, R., Aubinet, M., Bernhofer, C., Knohl, A., Laurila, T., Lindroth, A., Moors, E., Pilegaard, K., Seufert, G., 2007. Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites. *Geophys. Res. Lett.* 34, L01402, doi:10.1029/2006GL027880.
- Ruppert, J., Mauder, M., Thomas, C., Lueers, J., 2006. Innovative gap-filling strategy for annual sums of CO₂ net ecosystem exchange. *Agric. Forest Meteorol.* 138, 5–18.
- Stoy, P.C., Katul, G., Siqueira, M., Juang, J., Novick, K., McCarthy, H., Oishi, A., Oren, R., 2008. Role of vegetation in determining carbon sequestration along ecological succession in the southeastern United States. *Global Change Biol.* 14, 1–19.