Agricultural and Forest Meteorology xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Commentary

Reply to the comment on Vickers et al. (2009): 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 10 July 2009
Received in revised form 2 December 2009
Accepted 8 December 2009

Flux partitioning
Self-correlation
Spurious correlation

We thank Lasslop et al. (2010) for their remarks and interest, and appreciate the opportunity to clarify our manuscript. Hopefully this attention will increase awareness of the issue of self-correlation in the CO_2 flux community. There is currently a revival of the long-known concerns about self-correlation in the turbulence and boundary-layer community. Anderson (2009) suggests that because the degree to which self-correlation affects regression relationships is not simple to calculate in all cases, techniques that avoid the problem are preferred.

Self-correlation (Hicks, 1978; Klipp and Mahrt, 2004; Baas et al., 2006; Toba et al., 2008) has also been referred to as spurious correlation (Pearson, 1897; Kenney, 1982; Kenney, 1991; Jackson and Somers, 1991; Aldrich, 1995; Brett, 2004), spurious inference or faulty inference (Prairie and Bird, 1989) and as the shared variable problem in the statistics literature (see additional references in Vickers et al., 2009). 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. The degree of self-correlation is proportional to the ratio of the variances of the shared to the unshared variable. The relationship between the orientation of the self-correlation and of the true physical correlation is important. Depending on this relationship, self-correlation can be large or negligible. When the self-correlation is the same sign as the expected correlation and large enough, it can lead to false confidence in hypothesis testing. 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.

Vickers et al. (2009) evaluated the self-correlation between assimilation (GEP) and respiration (ER) caused by the shared variable problem that arises when using the standard method of partitioning the measured net CO_2 flux (NEE) from eddy-covariance measurements into its two components: GEP and ER. The standard method fits the temperature dependence of ER (nocturnal NEE) using night-time measurements only (when GEP is zero) and then applies the model to daytime NEE data. The shared variable problem arises because the estimates of GEP (=NEE-ER) and ER have a shared explicit dependence on ER. This is because GEP is not measured but instead is calculated as a residual to balance the budget. We showed that strong correlation can be found between GEP and ER even when the measurements of NEE and the model of ER are replaced with their random permutations (Fig. 1). Any relationships or correlations found based on the random permutations have no physical meaning and can only be due to the shared variable. This is what we and others have called self-correlation.

The important point is that random uncorrelated permutations of the observations can lead to a strong correlation between GEP and ER. This naturally raises the question: does the correlation significantly increase when using the real observations instead of the artificial data? If the answer is no (e.g., Fig. 1), we contend that a real correlation cannot be demonstrated based on the observations and the flux partitioning method. In such a case, most of the observed correlation is meaningless, even though it may be large, and even though a real correlation of the same sign is expected based on eco-physiological arguments. For the forest site we studied using our Method II day-night partitioning, we found that approximately one-half of the observed variance of GEP explained by ER was due to the shared variable problem.

The comment by Lasslop et al. argues that the approach used by Kenney (1982); Klipp and Mahrt (2004); Vickers et al. (2009) and others (see references in Vickers et al., 2009), over-estimates the self-correlation due to the shared variable problem because NEE is not an "original" variable, but rather consists of two distinct

---

[*] 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 Published by Elsevier B.V.
doi:10.1016/j.agrformet.2009.12.002

Please cite this article in press as: Vickers, D., et al., Reply to the comment on Vickers et al. (2009): Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2009.12.002

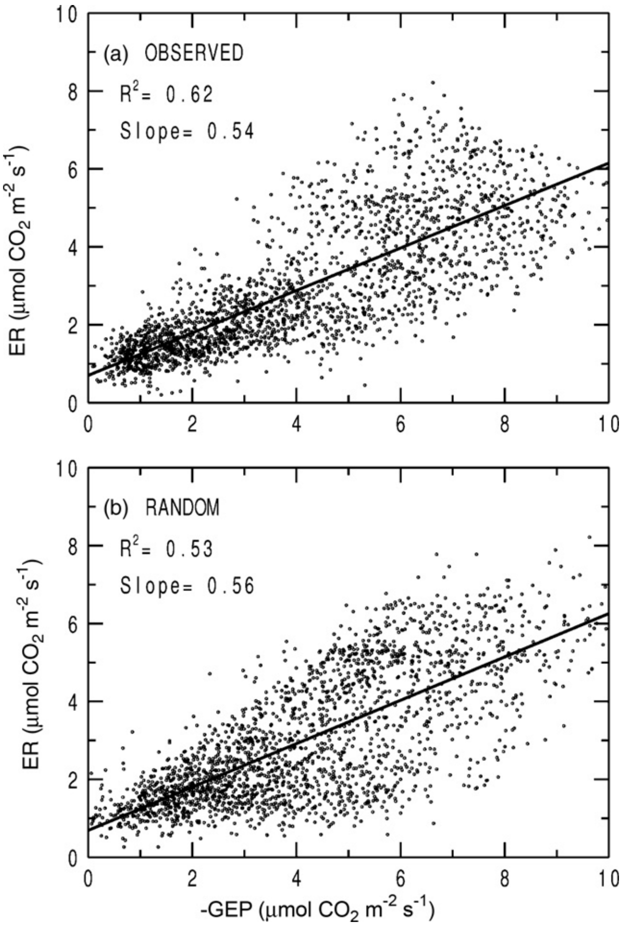

Fig. 1. Daily-averaged respiration (ER) as a function of assimilation (-GEP) for (a) the real observed NEE data and (b) one realization of the random permutations. The lower panel shows strong correlation even when replacing the real observations with their random permutations. The correlation in the lower panel has no physical meaning and is what we refer to as self-correlation due to the shared variable problem.

components: Reco and GPP (ER and GEP, respectively using our nomenclature). We contend that NEE is the measured variable and GEP and ER are derived variables.

The comment states "... Reco is not part of GPP." This appears to be a contradiction since GPP is defined as NEE-Reco. Because GPP is calculated as a residual, it has a direct dependence on NEE and Reco, and GPP and Reco have a shared variable. The comment uses the same argument (flawed in our opinion) to explain why they feel that the correlation between liver weight and total body weight minus liver weight is not spurious as liver weight is "not part of" total body weight minus liver weight.

The comment states that we over-estimate the self-correlation because we use the whole variance of ER. We can find no justification for this argument in their comment or in the literature. If our methodology overestimates the self-correlation, as suggested by the comment, that leads to the highly counter intuitive result that random permutations of the observations (artificial data) can lead to "real" (meaningful) correlation. We contend that because the random permutations no longer retain any real connections between GEP and ER, the correlation computed from such series has no physical meaning.

We agree that the variance of NEE will be reduced as the covariance between GEP and ER increases if the other terms are held constant, and this is what one would anticipate when thinking in eco-physiological terms, however, we do not see how this relates to the shared variable problem. In their Eq. (1) they have not accounted for the fact that GEP is not independent, but is defined as a residual (NEE-ER). The comment describes their experiment that generated artificial data where GEP and ER are normally distributed random variables with varying variances of GEP and varying correlation between the two. Based on this experiment, they concluded that the correlation between GEP and ER is completely real. We agree that the imposed correlation can be considered "real" in this case but we fail to see the relevance of this different problem to the one we have discussed. When there is no shared variable, there is no self-correlation due to the shared variable problem. With no self-correlation, all the observed correlation is real.

The comment states that our method "shows high values for the artificial correlation even if GPP and Reco are in fact uncorrelated". We do not understand this comment. We calculate the artificial correlation as the correlation between GEP and ER when using fabricated NEE data. If GEP and ER are uncorrelated based on the fabricated NEE, then the artificial correlation is zero.

Regarding Eqs. (2–4) in the comment, measurement errors are not a necessary condition for self-correlation as we and others have defined it based on the shared variable problem. However, there is similarity between our artificial data experiment and imposing measurement error. This is a distinction that may be open to interpretation. Baas et al. (2006) suggested that the impact of self-correlation can be investigated by imposing errors on the common variables. In such a case, the magnitude of the self-correlation will depend on the details of how the errors are imposed. Baas et al. (2006) referred to our artificial data approach as an alternate method for examining the self-correlation.

A possible solution for self-correlation in the CO_2 flux partitioning example is to combine independent estimates of ER and GEP. However, we contend that there is no reliable way to measure GEP using eddy-covariance measurements, and that therefore self-correlation between GEP and ER is unavoidable when dealing with the observations. The situation is less clear when models or combinations of models and observations are employed to describe the fluxes, where the self-correlation will depend on the model details. If both of the CO_2 flux components and the net flux could be measured, then the imbalance in the budget could be studied, as for example with the surface heat budget, where net radiation drives sensible heat flux, latent heat flux and soil heat flux. In the surface heat budget, all the terms can be directly measured and are independent, and there is no self-correlation. However, in the CO_2 budget, GEP is calculated as a residual to balance the budget and thus has a direct imposed dependence on the other terms in the budget.

Although not mentioned in their comment, we note that our estimates of the real correlation (see Tables 1–2 in the manuscript) are not rigorously determined. Following Klipp and Mahrt (2004) and others, we take the difference between the variance explained by the real observations and the variance explained by random permutations of the observations as an estimate of the variance explained by real underlying processes.

As we stated in Vickers et al. (2009) (see abstract and conclusions), we do not argue against a true strong correlation between GEP and ER based on eco-physiological considerations. We do argue that one cannot quantify the true correlation using eddy-covariance flux measurements and standard flux partitioning methods without accounting for the self-correlation that arises from the shared variable problem.

Acknowledgments

We thank Larry Mahrt of COAS and Alix Gitelman of the Statistics Department at Oregon State University for helpful discussions.

Please cite this article in press as: Vickers, D., et al., Reply to the comment on Vickers et al. (2009): Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2009.12.002

References

Aldrich, J., 1995. Correlations genuine and spurious in Pearson in Yule. Stat. Sci. 10, 364–376.

Anderson, P.S., 2009. Measurement of Prandtl number as a function of Richardson number avoiding self-correlation. Boundary-Layer Meteor. 131, 345–362.

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.

Brett, M.T., 2004. When is correlation between non-independent variables "spurious"? Oikos 105, 647–656.

Hicks, B.B., 1978. Some limitations of dimensional analysis and power laws. Boundary-Layer Meteor. 14, 567–569.

Jackson, D.A., Somers, K.M., 1991. The spectre of spurious correlations. Oecologia 86, 147–151.

Kenney, B.C., 1982. Beware of spurious self-correlation! Water Resour. Res. 18, 1041–1048.

Kenney, 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.

Klipp, C.L., Mahrt, L., 2004. Flux–gradient relationship, self-correlation and intermittency in the stable boundary layer. Quart. J. R. Meteor. Soc. 130, 2087–2104.

Lasslop, G., Reichstein, M., Detto, M., Richardson, A., Baldocchi, D., 2010. Comment on Vickers et al.: self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agric. Forest Meteor., this volume.

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.

Prairie, Y.T., Bird, D.F., 1989. Some misconceptions about the spurious correlation problem in the ecological literature. Oecologia 81, 285–288.

Toba, Y., Suzuki, N., Komori, S., 2008. Discrimination of spurious self-correlation in nondimensionalized analyses of fluid dynamical data. J. Oceanogr. 64, 393–397.

Vickers, D., Thomas, C., Martin, J., Law, B., 2009. Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agric. Forest Meteor. 149, 1552–1555.

Please cite this article in press as: Vickers, D., et al., Reply to the comment on Vickers et al. (2009): Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2009.12.002