Commentary

Comment on Vickers et al.: Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO₂ fluxes

Gitta Lasslopa,*, Markus Reichstein a, Matteo Detto b, Andrew D. Richardson c, Dennis D. Baldocchi b

a Max-Planck Institute for Biogeochemistry, Model-Data Integration Group Postfach 10 01 64, 07701 Jena, Germany
b Ecosystem Sciences Division, Department of Environmental Science, Policy and Management, University of California, 137 Mulford Hall, Berkeley, CA 94720, USA
c Department of Organismic and Evolutionary Biology, Harvard University, HUH 22 Divinity Ave, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 15 June 2009
Received in revised form 3 November 2009
Accepted 4 November 2009

Keywords:
Flux partitioning
Spurious self-correlation
Eddy-covariance

The paper of Vickers et al. (2009) raises the important issue of self-correlation, also named spurious correlation or artificial correlation, in the context of partitioning measured eddy-covariance (EC) CO₂ fluxes. Vickers et al. argue that there is a spurious correlation between primary production (GPP) and ecosystem respiration (Reco) when these component fluxes are jointly estimated from the measured net ecosystem exchange (NEE) of CO₂. Spurious correlations can arise, for instance, when an independent variable x is compared with a dependent variable y that is computed from x as x = y (Kenney, 1982). In this example x and y are the “original” variables and z is derived from these. The higher the variance of the shared variable x compared to the variance of y, the higher the spurious correlation between x and z (Kenney, 1982).

Standard flux partitioning methods derive GPP and Reco (here, by definition, both positive quantities, whereas negative NEE indicates net uptake by terrestrial vegetation) by fitting a respiration model to nighttime EC data (where GPP = 0, as there is no photosynthesis in the dark), extrapolating the model into daytime, and deriving the estimated GPP as Reco-NEE (Desai et al., 2008; Reichstein et al., 2005). Here, Reco corresponds to x, NEE to y, and GPP to z = x − y. This description can be further split up as described in Vickers et al. (2009) method II: when using daily sums, or data further aggregated, the data can be split into their nighttime and daytime part then GPP = NEE_{\text{day}} − Reco_{\text{day}}. The correlation between GPP and Reco could be expected to be spurious, as the equation to derive GPP contains Reco or at least the daytime part. Even when splitting the variables into day and nighttime, the estimate of the spurious correlation is not small (Vickers et al., 2009).

A key difference between the flux partitioning and the Kenney (1982) example is that NEE is not an “original” independent random variable but rather consists of two distinct components: Reco and GPP. With the operation Reco − NEE the influence of Reco is removed from NEE thus Reco is not part of GPP. But note that any error in the estimation of Reco would be directly transferred to GPP. Spurious correlation is often illustrated with the example of plotting the weight of the liver against total body weight. Variability of the liver directly causes a variability of the total body weight and thus a correlation (Kainz et al., 2009). In case of the flux partitioning Reco is the analogue of the liver, but NEE is total body weight and GPP is total body weight minus liver. The correlation between liver and total body weight minus liver would not be expected to be spurious (unless there were measurement errors), as the liver is not part of the second variable and does not contribute to its variability.

Here we are not arguing against the general importance and the necessity to be aware of the problem of self-correlation when estimating GPP and Reco from the same EC data, but self-correlation is important because of the error in Reco rather than because Reco being a shared variable. We show that the spurious correlation estimated using the whole variance of Reco with the

Fig. 1. Standard deviation of NEE versus the correlation (R) between GPP and Reco.
The “artificial correlation” between (Fig. 1). For this constructed dataset our understanding of the varying correlation between the two, distributed random variables with varying variances of although exchange is by definition the difference between correlation between NEE (Fig. 2). Vickers et al. (2009), which would predict a high spurious identity implies that the variance (var) of NEE decreases with increasing covariance (covar) between GPP and Reco, since

\[
\text{var}(\text{NEE}) = \text{var}(\text{GPP}) + \text{var}(\text{Reco}) - 2\text{covar}(\text{GPP, Reco})
\]  

(1)

This contradicts the concept of spurious correlation following Eq. (2) in Vickers et al. (2009), which would predict a high spurious correlation for a low variance of NEE. Note that this equation is based on the positive definition of GPP and Reco. Following the micrometeorological sign convention, where a flux towards the surface is negative, a negative covariance would be added resulting in the same decreased variance of NEE. We illustrated this relationship using artificial data where GPP and Reco are normally distributed random variables with varying variances of GPP and varying correlation between the two, NEE is computed as Reco-GPP (Fig. 1). For this constructed dataset our understanding of the correlation between GPP and Reco is that, it is completely real and the “artificial correlation” between GPP and Reco should be zero, although GPP can be computed as Reco-NEE.

Two approaches have been proposed to estimate the self-correlation: (1) based on the variances of the two original variables (x and y) (Kenney, 1982) and (2) a method that decorrelates NEE and Reco (Vickers et al., 2009). GPP is then computed with the randomized Reco, assuming that Reco is part of GPP, the correlation between GPP and the randomized Reco is the self-correlation. They both show the same high values for the “artificial correlation” even if GPP and Reco are in fact uncorrelated (Fig. 2a). Using the whole variance of Reco to estimate the spurious correlation can lead to a strong overestimation of spurious correlation and a misleading estimate of the real correlation, if the real correlation is estimated as observed minus spurious correlation as in Vickers et al. (2009) (Fig. 2b).

So far we assumed perfect observations of NEE and perfect estimates of Reco, which cannot be expected (Desai et al., 2008; Lasslop et al., in press; Richardson et al., 2006). Including the errors the equations can be changed into

\[
\text{NEE}_{\text{obs}} = \text{GPP}_{\text{true}} + \text{Reco}_{\text{true}} + \varepsilon_{\text{NEE}}.
\]

(2)

\[
\text{Reco}_{\text{est}} = \text{Reco}_{\text{true}} + \varepsilon_{\text{reco}}
\]

(3)

\[
\text{GPP}_{\text{est}} = \text{Reco}_{\text{est}} - \text{NEE}_{\text{obs}} = \text{GPP}_{\text{true}} + \varepsilon_{\text{reco}} + \varepsilon_{\text{NEE}}.
\]

(4)

where \( \varepsilon \) denote error terms, the subscript “true” denotes the true value of the system, “obs” and “est” the observed and estimated values, respectively. Thus spurious correlations can be introduced via error terms that are part of both, Reco and GPP, but only if they contribute to the variance, e.g. they are not constant. This explains the high estimates of self-correlation when using the randomized time series of Reco to compute GPP, as the relation between NEE and Reco is destroyed, the Reco error is maximized by the randomization and \( \text{Reco}_{\text{true}} = 0 \).

To derive relationships between Reco and GPP without having to face the problem of spurious correlations or to explore the extent of a spurious correlation we recommend to combine independent estimates for such analysis. For instance, respiration can be derived from chamber measurements and combined with GPP from eddy-covariance measurements. If only eddy-covariance measurements are available Reco can be derived using only nighttime data and GPP using only daytime data. The influence of the error in Reco on the correlation between GPP and Reco as well as additional evidence of real correlation between GPP and Reco is further explored in a forthcoming manuscript (Lasslop et al., in press).

Acknowledgements

We would like to thank S. Pohl, Department of Methodology and Evaluation research, University of Jena, C. Strobl, Department of Statistics, LMU München and C. Middelhoff (former member of the MPI for Biogeochemistry, Jena) for valuable discussion on the topic. We acknowledge the comments of an anonymous reviewer that helped us to improve the clarity of the manuscript.

References


