Agricultural and Forest Meteorology 232 (2017) 635–649

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Preserving the variance in imputed eddy-covariance measurements: Alternative methods for defensible gap filling

Sujit Kunwor[a], Gregory Starr[a], Henry W. Loescher[b,c], Christina L. Staudhammer[a,*]

[a] University of Alabama, Department of Biological Sciences, Tuscaloosa, AL 35487, USA
[b] Battelle – National Ecological Observatory Network Inc., Boulder, CO 80301, USA
[c] Institute of Alpine and Arctic Research (INSTAAR), University of Colorado, Boulder, CO 80301, USA

A R T I C L E I N F O

Article history:
Received 6 February 2016
Received in revised form 7 September 2016
Accepted 24 October 2016
Available online 29 October 2016

Keywords:
Eddy covariance
Net ecosystem exchange
Gap-filling
Variance preservation
Parameter prediction
Orthonormal wavelet transformation

A B S T R A C T

The high utility of eddy covariance (EC) data has made it the cornerstone of carbon dynamics research for more than two decades. However, a substantial number of measurements from EC data can be missing for various reasons. Robust gap-filling methods are required to estimate carbon budgets from net ecosystem exchange measurements of CO_2 (NEE) with high precision and accuracy. While the gap-filled methods used have provided unbiased estimates of annual NEE, little research has been done on preserving the variance structures associated with gap-filled flux data. In this project, we used EC data from a longleaf pine ecosystem located in the southeast US to investigate variance preservation in gap-filling methods.

We used three non-linear regression approaches to impute artificially created gaps of different sizes via light and temperature response curves: 1) "traditional" fixed monthly window, 2) moving window, and 3) moving window with parameter prediction using physiological drivers. The results of gap-filling simulations showed that the variability of NEE estimates made with moving window and parameter prediction methods were closer to that of observed NEE, whereas the traditional method had overall lower variability. The average root mean square errors (RMSE) of predictions was lower for moving window and parameter prediction (3.38 and 3.22, respectively), versus that of the traditional method (3.42) over one year, including both daytime and nighttime data across all gap sizes. The variances associated with moving window and parameter prediction methods were 52% and 57%, respectively, of the observed variance, versus that of the traditional (51%), while the average of first-order autocorrelation coefficients was 0.76 for each method compared to 0.58 for observed. The results showed that the moving window approaches provided better estimates (lower RMSEs and more similar variance) at annual scales, yet underestimated the observed variance. These results contribute toward the development of a framework of standardized gap-filling approaches which maintain variation inherent in EC data. Moreover, these results call for further research on potential environmental drivers and their interactions for inclusion in gap-filling models, as well as exploration of sampling size of estimation windows and averaging time (half hour) of flux data to promote variance maintenance and decrease the autocorrelation of predictions.

1. Introduction

The study of terrestrial carbon dynamics plays a crucial role in understanding the biological processes that contribute towards atmospheric CO_2 concentrations (Baldocchi et al., 1996; Goulden et al., 1996; Papale et al., 2006). In the present context of global climate change, the eddy covariance (EC) method has improved our understanding of temporally and spatially integrated net ecosystem exchange of CO_2 (NEE) (Baldocchi et al., 1996; Schimel, 2000; Schimel et al., 2001; Stoy et al., 2009). The data produced by the EC method is not only useful to estimate ecosystem-level annual carbon budgets, *i.e.*, temporal integrals of NEE, but also helps to establish functional relationships between NEE and micrometeorological variables, and to devolve below-canopy ecological processes (Loescher et al., 2003; Sierra et al., 2011).

Despite the mathematical and theoretical foundations of the EC method, data cannot be collected under all climatic conditions (Burba, 2013). Datasets routinely have missing data, which arise during the data collection period or from filtering out data that do not meet micrometeorological assumptions (Goulden et al., 1996; Gu et al., 2005; Hollinger and Richardson, 2005; Papale et al., 2006). The EC datasets often have missing periods as a result of: i) rain and

* Corresponding author at: P.O. Box 870344, Tuscaloosa AL 35487, USA.
E-mail address: cstaudhammer@ua.edu (C.L. Staudhammer).

http://dx.doi.org/10.1016/j.agrformet.2016.10.018
0168-1923/

condensation in the sampling path, ii) incomplete datasets at the desired averaging times (*i.e.*, 30-min) during system calibration or maintenance, iii) poor coupling of the canopy with the above atmospheric conditions, as defined by the friction velocity, u*, or iv) excessive variation from the half-hourly mean based on an analysis of standard deviations for wind and CO_2 statistics (Whelan et al., 2013). The series of missing data may extend over hours or days, and are non-randomly distributed (Hagen et al., 2006). For example, at an EC site on the Coastal Plain in the southeastern US, missing data accounted for 63%, 52% and 66% of nighttime values for mesic, intermediate and xeric longleaf pine savannah ecosystems, respectively (Whelan et al., 2013; Starr et al., 2015). EC data gaps are commonly filled utilizing various empirical or semi-empirical methods (Falge et al., 2001; Moffat et al., 2007).

Communities of long-term carbon measurement such as FLUXNET (Baldocchi et al., 2001), AmeriFlux, and the Integrated Carbon Observation System (ICOS) have stressed standardized and statistically defensible gap-filling approaches in order to make synthesis activities feasible, primarily for annual carbon budgets. Despite this effort, a number of gap-filling approaches are used and include non-linear regression (NLR), look-up tables (LUT), mean diurnal variation methods (MDV) (*e.g.*, Falge et al., 2001), artificial neural networks (ANN) (*e.g.*, Papale and Valentini, 2003; Braswell et al., 2005; Moffat et al., 2007), Marginal Distribution Sampling (MDS) (*e.g.*, Reichstein et al., 2005), data assimilation (DAM) or Bayesian model approaches (Gove and Hollinger, 2006). However, their performances vary over time and scale (Hollinger et al., 2004; Moffat et al., 2007), and may not always be statistically defensible. The non-random nature of missing data also introduces bias into predictions, and the autocorrelation among observations artificially deflates their standard errors (Anderson, 1954).

A study of a few of these gap-filling techniques (Falge et al., 2001) concluded that MDV, NLR and LUT have similar performance in terms of the systematic bias that is introduced in annual sums of NEE. Furthermore, this bias was found to be proportional to the percentage of gaps filled. While their methods included light and temperature relationships to fill NEE, they recommended adding additional predictors, such as vapor pressure deficit and soil water content, to reduce this systematic bias.

Moffat et al. (2007) followed on this work with a comparison of common gap-filling methods applied to a wider range of datasets that included artificially created gaps of varying lengths in time. This study also showed less bias of gap-filling approaches using additional new techniques such as ANN. However, studies thus far have not adequately addressed preserving the variance in gap-filled data as that observed from the measured NEE.

The variance in data from environmental drivers (*e.g.*, air temperature) used in gap-filling methods has been shown to amplify or dampen flux estimates depending on the degree of convexity in the flux-temperature relationship (Sierra et al., 2011). Models that rely on average temperature underestimate flux values compared to those that preserve observed variance as a consequence of Jensen's Inequality, and therefore long-term estimates of NEE will be further biased (Moffat et al., 2007). This logic would extend to any case using flux data to estimate any non-linear behavior, as would be expected with changing chronic disturbance (Smith et al., 2009). Hence, additional development of techniques is needed that advances our ability to (i) preserve the variance structures inherent in EC data, and (ii) reduce the systematic bias as well as the error due to gap-filling techniques.

End-to-end systematic and random error propagation from EC data collection and processing steps has not been ensembled into a defensible uncertainty budget, *i.e.*, classic metrology (JCGM, 2008; ISO, 1995). Instead, the state-of-the-science has been assessing individual sources of both systematic and random errors. For example, observations before gap-filling have systematic and random errors (Goulden et al., 1996; Loescher et al., 2006a), which can be further propagated during gap-filling practices. This issue has been examined by a few studies. Hollinger and Richardson (2005) compared flux measurements from two adjacent tower fluxes to estimate random errors and showed double exponential distribution and heteroscedastic variances. Uncertainty budgets in calibrations and random sources of error were estimated from the AmeriFlux network (Ocheltree and Loescher, 2007), and overall uncertainties in EC techniques can be found in Loescher et al. (2006a). Moffat et al. (2007) compared the error contribution by different gap-filling methods. Wang et al. (2015) used a residual bootstrap approach to quantify uncertainty associated with random measurement and gap-filling errors. The limitations currently found in gap filling techniques establish a need to further investigate gap filling techniques that are more robust, statistically defensible, and work toward preserving the native variance. In this study, we develop statistically defensible methods to re-introduce the variance structure in gap-filled NEE data, and compare and contrast the results of three model approaches. It is not the intention of this study to develop uncertainty budgets.

We evaluated the ability of gap-filling methods to preserve variance inherent in the data structure as a means to provide more robust estimates of annual sums of NEE. Three different gap-filling approaches based on non-linear light and temperature response models were compared. First, a traditional method utilizing static monthly data was used to parameterize these models. In the second method, the same models were used with a monthly moving window rather than the static window. The third method used models with a monthly moving window and the incorporation of additional micrometeorological measurements as additional explanatory variables to better account for the controls on NEE variability. We hypothesized that the gap-filling method with moving window (method 2) would better capture variance in NEE by the use of varying parameters estimated from moving the sampling window. We also hypothesized that using additional environmental variables in the model (method 3) would improve the variance preservation of NEE from 30-min to annual time scales.

2. Methods

After EC data were collected from the field and processed into 30-min estimates (rf. Whelan et al., 2015; Starr et al., 2015, 2016), our first step was to choose a suitable base dataset for analyses. From this base dataset, four case study datasets were created for testing different gap-filling methods. While two of our methods could be directly tested with the case study datasets, implementation of our third method required additional preliminary analyses using the base dataset (Fig. 1).

2.1. Data collection and processing

Carbon fluxes and micrometeorological data were obtained from three established sites along an edaphic gradient (mesic, intermediate and xeric) at the Joseph Jones Ecological Research Center (JJERC) in southwestern Georgia, USA (31.2201°N, 84.4792°W). NEE has been measured continuously at all three sites starting in October 2008 with open path EC techniques, which includes an infrared gas analyzers (IRGA, LI-7500, LICOR Inc., Lincoln, NE), and 3-D sonic anemometers (CSAT-3, Campbell Scientific Inc., Logan, UT). The instruments were installed in towers above the forest canopy at heights: 34.4, 37.5 and 34.9 m for mesic, intermediate and xeric sites, respectively (~4.0 m above mean canopy height). Micrometeorological data were collected and stored in a datalogger (CR5000, Campbell Scientific Inc., Logan, UT).

Raw EC data were processed following methods of Starr et al. (2016) and Whelan et al. (2013) using EdiRe (v.1.4.3.1184; Clement