Atmospheric Environment 101 (2015) 312–327

Contents lists available at ScienceDirect

Atmospheric Environment

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

Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling

Olaf Menzer [a,*], Wendy Meiring [b], Phaedon C. Kyriakidis [a], Joseph P. McFadden [a]

[a] Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA
[b] Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA

HIGHLIGHTS

- Estimates of three years of carbon dioxide exchange in an urban environment.
- Gap-filling models explained between 64% and 88% of the variability in the data.
- Traffic is an important driving variable for gap-filling models at a suburban site.
- Modeled annual carbon budgets varied by a factor of two between wind sectors.
- Machine learning based gap-filling can be applied at other heterogeneous sites.

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form
1 November 2014
Accepted 3 November 2014
Available online 5 November 2014

Keywords:
Carbon dioxide
Eddy covariance
Gap-filling
Urban ecosystem
Machine learning
Uncertainty
Spatial heterogeneity

ABSTRACT

A small, but growing, number of flux towers in urban environments measure surface–atmospheric exchanges of carbon dioxide by the eddy covariance method. As in all eddy covariance studies, obtaining annual sums of urban CO_2 exchange requires imputation of data gaps due to low turbulence and non-stationary conditions, adverse weather, and instrument failures. Gap-filling approaches that are widely used for measurements from towers in natural vegetation are based on light and temperature response models. However, they do not account for key features of the urban environment including tower footprint heterogeneity and localized CO_2 sources. Here, we present a novel gap-filling modeling framework that uses machine learning to select explanatory variables, such as continuous traffic counts and temporal variables, and then constrains models separately for spatially classified subsets of the data. We applied the modeling framework to a three year time series of measurements from a tall broadcast tower in a suburban neighborhood of Minneapolis-Saint Paul, Minnesota, USA. The gap-filling performance was similar to that reported for natural measurement sites, explaining 64% to 88% of the variability in the fluxes. Simulated carbon budgets were in good agreement with an ecophysiological bottom-up study at the same site. Total annual carbon dioxide flux sums for the tower site ranged from 1064 to 1382 g C m^{-2} yr^{-1}, across different years and different gap-filling methods. Bias errors of annual sums resulting from gap-filling did not exceed 18 g C m^{-2} yr^{-1} and random uncertainties did not exceed $\pm$44 g C m^{-2} yr^{-1} (or $\pm$3.8% of the annual flux). Regardless of the gap-filling method used, the year-to-year differences in carbon exchange at this site were small. In contrast, the modeled annual sums of CO_2 exchange differed by a factor of two depending on wind direction. This indicated that the modeled time series captured the spatial variability in both the biogenic and anthropogenic CO_2 sources and sinks in a reproducible way. The gap-filling approach developed here may also be useful for inhomogeneous sites other than urban areas, such as logged forests or ecosystems under disturbance from fire or pests.

1. Introduction

Urban areas account for at least 70% of anthropogenic carbon dioxide emissions (Canadell et al., 2009), mainly due to the burning of fossil fuels for transportation, space heating, industrial uses, and electric power generation (which may take place outside cities, Satterthwaite (2008)). This will become only more important as an

* Corresponding author.
E-mail address: menzer@geog.ucsb.edu (O. Menzer).

http://dx.doi.org/10.1016/j.atmosenv.2014.11.006
1352-2310/

increasing proportion of the world's population lives in cities (Seto et al., 2012). Cities provide opportunities for emission reduction through urban design choices. Despite the dominant role of anthropogenic emissions in the urban environment, vegetated areas within cities have been shown to influence the net CO_2 emissions. Cities with a higher percentage of vegetation cover tend to have lower net CO_2 exchange per unit area (Velasco and Roth, 2010). CO_2 uptake by vegetation reduces daytime emissions, but does not transform sites that are net carbon sources into sinks (Grimmond et al., 2002). However, urban vegetation and soils have a significant effect on the seasonality of CO_2 exchange in temperate cities (Peters and McFadden, 2012). Yet photosynthetic uptake and respiratory release of CO_2 from urban green spaces are seldom included in carbon emission inventories of cities (Nordbo et al., 2012), which calculate carbon budgets by bottom-up modeling. Recently, Christen et al. (2011) modeled ecosystem uptake and emissions of CO_2 by combining chamber measurements of soil and above ground biomass respiration as well as irradiance measurements to infer photosynthetic activity and up-scale it to the neighborhood level using airborne light detection and ranging (LIDAR) and satellite data. Nevertheless, uncertainties remain among the different methods to quantify CO_2 fluxes of urban vegetation, and studies need to be extended to cities in different climate zones and the southern hemisphere (Weissert et al., 2014).

Seasonal and annual sums of CO_2 emissions in cities are important for planning and design at fine spatial scales such as at the neighborhood, block, or building level (Kellett et al., 2013). The eddy covariance (EC) method (Baldocchi, 2008) has been used to validate local scale emission inventories (Christen et al., 2011), study the impact of urbanization on CO_2 fluxes (Ramamurthy and Pardyjak, 2011), and improve the understanding of urban ecosystems (Velasco and Roth, 2010). EC provides a means to directly measure the net CO_2 flux, thereby accounting for both the dominant anthropogenic sources of CO_2 emissions and the contributions of vegetation and soils (Feigenwinter et al., 2012). The limitations of EC measurements include fragmentation of data sets due to system failures (caused by e.g., snow, lightning, or birds) and low turbulence atmospheric conditions (e.g., during night), the latter resulting in rejection of some observations. On an annual basis, these so-called *gaps* typically account for 20–60% of a flux data set (Moffat et al., 2007). Gap-filling of the flux time series is essential to obtain annual sums of net CO_2 flux which can be used to calculate carbon budgets and to evaluate process-based model predictions. If gaps are filled according to the respective wind direction, spatially unbiased estimates of CO_2 flux that do not depend on the frequency distribution of the wind rose at the site can be provided. This could potentially help to quantify the location bias of individual tower sites. In addition, implementing meaningful gap-filling offers insights into the controls of urban carbon fluxes and a subsequent partitioning of the net flux into different components.

Eddy covariance time series of CO_2 exchange are routinely gap-filled using methods such as look-up tables (e.g., Falge et al. (2001); Reichstein et al. (2005)), nonlinear regressions (e.g., Falge et al. (2001); Hollinger et al. (2004)) and artificial neural networks (ANN, e.g., Papale and Valentini (2003); Braswell et al. (2005); Moffat (2012)). All of these are based on well understood relationships of the fluxes with environmental drivers such as temperature and light. Typically, for sites in natural ecosystems, these gap-filling methods can explain 60–90% of the variability of daytime CO_2 fluxes and 25–66% of the variability of nighttime CO_2 fluxes (Moffat et al., 2007).

In the urban environment, however, gap-filling of EC measurements is more complex compared to natural and managed ecosystems, mainly due to two reasons. First, there are additional explanatory variables related to anthropogenic CO_2 emissions, such as vehicular traffic or the diurnal pattern of building energy use. Second, urban surfaces are more heterogeneous in terms of the spatial pattern of vegetation and localized sources of CO_2 emission such as buildings and motor vehicles (Vesala et al., 2008; Pickett et al., 2008; Kotthaus and Grimmond, 2012). This means that urban CO_2 flux measurements have a higher dependence on the prevailing wind direction (Vesala et al., 2008; Kordowski and Kuttler, 2010; Crawford et al., 2011; Järvi et al., 2012). The underlying patterns of anthropogenic and ecological sources and sinks of CO_2 differ among urban sites (Velasco and Roth, 2010; Crawford et al., 2011; Nordbo et al., 2012), making it a challenge to identify and rank the importance of explanatory variables needed to model the net CO_2 exchanges in cities.

Most of the urban flux studies in the literature have circumvented the problem of gap-filling by working with diurnal fluxes averaged over monthly time spans (Grimmond et al., 2002; Soegaard and Moller-Jensen, 2003; Christen et al., 2011; Peters and McFadden, 2012), which is sufficient for evaluation of models or for comparison to satellite data with a coarser temporal resolution. Some studies have taken advantage of site-specific conditions such as the fraction of vegetation in the footprint of a tower in Vancouver, Canada (Crawford et al., 2011), and Crawford and Christen (2014) modelled fluxes based on surface fractions in the changing turbulent source area. Other work on gap-filling specific to urban areas was carried out by Schmidt et al. (2008), who used artificial neural networks (ANNs) and radial basis function networks (RBFs) to gap-fill fluxes measured over a summer season in Münster, Germany, and by Kordowski and Kuttler (2010) for longer time series in Essen, Germany. More recently, Järvi et al. (2012) reported a gap-filled five-year data set of CO_2 fluxes in Helsinki, Finland, using ANNs and a mean diurnal variation approach.

Here, we gap-filled a three year time series of carbon fluxes from a tall radio broadcast tower in a first-ring suburb of Minneapolis-Saint Paul, Minnesota, USA, using a wide selection of auxiliary measurements. Our key objectives were to: 1) identify the controls of CO_2 emissions specific to the urban environment and how they varied in time and space (throughout the tower footprint); 2) train machine learning regression models that can reproduce fluxes measured in spatially heterogeneous landscapes, and evaluate model performance and error; 3) assess the heterogeneity of the flux source area using various potential explanatory variables and determine their importance for modeling and gap-filling of EC measurements in the urban environment; 4) calculate seasonal and annual CO_2 flux sums for the tower site observed flux footprint and for the main land use types within it, as well as provide estimates of associated systematic and random uncertainties.

2. Materials and methods

2.1. Study site

Our study was conducted in a first-ring suburban neighborhood immediately outside the city of Saint Paul, Minnesota, USA (44° 59′ N, 93.1° 11′ W). The neighborhood has approximately 1000 inhabitants km^{-2} and a housing density of 350 housing units km^{-2} (Radeloff et al., 2005). As described in more detail in Peters et al. (2011), the site was characterized by a cold temperate climate zone influenced by the urban heat island effect (Todhunter, 1996; Sen Roy and Yuan, 2009). The relatively flat terrain made this landscape suitable for eddy covariance measurements from the 150-m tall KUOM broadcast tower (Fig. 1). The average tall tower flux footprint area at the 40 m level had 82% vegetation cover, which consisted primarily of open turfgrass lawns, forested patches, and isolated trees with a mean tree height of 12 m. The land-use types within the tower footprint included