Generating Spatially Robust Carbon Budgets From Flux Tower Observations

Article *in* Geophysical Research Letters · January 2020

DOI: 10.1029/2019GL085942

5 authors, including:

Anne Griebel
Western Sydney University

Elise Pendall
Western Sydney University

George Burba
LI-COR Biosciences

Stefan Metzger
National Ecological Observatory Network

Generating spatially robust carbon budgets from flux tower observations

Running head: Spatially robust carbon budgets

Anne Griebel[a,*], Daniel Metzen[a], Elise Pendall[a], George Burba[b,c], and Stefan Metzger[d,e]

[a]Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith, NSW 2571, Australia

[b]LI-COR Biosciences, 4647 Superior St, Lincoln, Nebraska, USA

[c]R. B. Daugherty Water for Food Global Institute, School of Natural Resources, University of Nebraska, Lincoln, Nebraska 68583, USA

[d]National Ecological Observatory Network Program, Battelle, Boulder, Colorado, USA

[e]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

* Corresponding author: Anne Griebel

email: a.griebel@westernsydney.edu.au

phone: +61 2 4570 1668

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019GLO85942

Key points

- Variations in spatial and temporal sampling patterns over heterogeneous sites introduce variability in carbon budgets from flux towers.
- Spatial heterogeneity is prevalent across the network and results in up to 60% overestimation of mean biome productivity.
- A new method is proposed to account for contributions of sampling variations to carbon budget estimates at heterogeneous flux sites.

Abstract

Estimating global terrestrial productivity is typically achieved by rescaling individual flux tower measurements, traditionally assumed to represent homogeneous areas, using gridded remote sensing and climate data. Using 154 locations from the FLUXNET2015 data base, we demonstrate that variations in spatial homogeneity and non-uniform sampling patterns introduce variability in carbon budget estimates that propagate to the biome scale. We propose a practical solution to quantify the variability of vegetation characteristics and uniformity of sampling patterns, and moreover, account for contributions of sampling variations over heterogeneous surfaces to carbon budgets from flux towers. Our proposed space-time-equitable budgets reduce uncertainty related to heterogeneities, allow for more accurate attribution of physiological variations in productivity trends and provide more representative grid cell averages for linking fluxes with gridded data products.

Plain language summary

Uniform vegetation characteristics are a core requirement for atmospheric measurements of carbon, water and energy fluxes using monitoring stations called flux towers. The variability of ecosystem properties was commonly assumed to be negligible because no standard approach existed to quantify it. We developed a standardized approach to quantify the variability of vegetation properties from satellite data, and applied it to 154 sites within the global network of flux towers. We show that non-uniform vegetation characteristics are prevalent across the network, and that sampling patterns at such sites influence carbon budget estimates at individual flux sites. This artefact propagates when flux tower data is used to extrapolate plant productivity to larger scales. We propose a practical solution that accounts for the sampling variability over sites with variable vegetation characteristics, which removes this previously unaccounted artefact from a sites annual carbon budget. Our

spatially representative budgets bridge a critical gap in scaling ecosystem processes from individual sites to the globe and allow for more accurate estimates of terrestrial carbon uptake.

Keywords: heterogeneity, carbon budget, scaling, uncertainty, footprint, eddy covariance

1 Introduction

Ecosystem-scale carbon sequestration rates from flux towers present our current best means to directly quantify carbon cycling from the "bottom-up". Observations from individual sites are contributed to local, regional and global networks that strongly enhance the value of single site data (Baldocchi 2008). While scientific motivation, funding and data processing vary at a site-level (Novick et al. 2018), these distributed observations are aggregated in a standardized global network collection (FluxNet). This allows researchers to decipher similarities and differences among individual sites and biomes, and to rescale local observations to regional and continental scales. Such synthesis activities, along with related remote sensing and modelling efforts, form an integral foundation for global assessments of climate change impact (e.g., IPCC Reports) which informs the general public on the state of climate change.

Regional and global scale estimates of carbon cycling depend on appropriate aggregation of individual tower fluxes, grid-scale observations of vegetation from repeat satellite observations, and Earth system modeling (e.g. Ahlström et al. 2015; Bodesheim, Jung, Gans, Mahecha, and Reichstein 2018; Chen et al. 2017; Huang et al. 2018; Jung et al. 2011). An essential prerequisite to justify upscaling from sites to the globe from bottom-up measurements requires homogeneous site-level characteristics (i.e. topography, soil and vegetation properties; Chen et al. 2012; Kim et al. 2006). Spatial homogeneity is defined as "the measurement that is taken in the space-time domain reflects the actual conditions in the same or different space-time domain taken on a scale appropriate for a specific application" (Nappo et al. 1982); that is in a perfectly homogeneous ecosystem, the magnitude of fluxes would be uniform with wind direction as individual tower locations would provide spatially representative measurements. However, site-level characteristics are in practice rarely completely homogeneous (Foken and Leclerc 2004; Giannico et al. 2018; Goeckede et al. 2008; Schmid and Lloyd 1999; Stoy et al. 2013), and changes in the size of observed fluxes may result from changes in the prevailing wind directions that cause systematic differences in the area that is being sampled rather than from changes in ecological activity (Fig. 1).

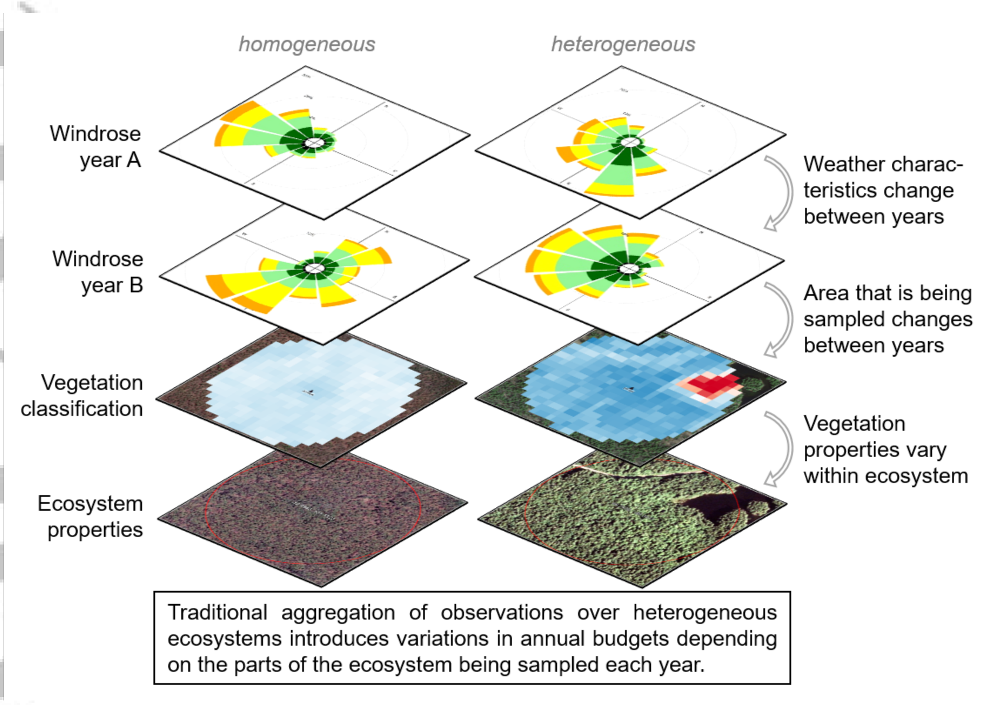

Figure 1: Conceptual illustration outlining how sampling variations can introduce variability in annual carbon budgets at sites with heterogeneous surface characteristics. This variability in carbon fluxes due to the area being sampled is unaccounted for in traditional carbon budgets, which simply aggregate observations over time. Our proposed solution accounts for wind-induced temporal and spatial sampling variations over heterogeneous ecosystems (differing annual windroses; upper two rows). We use variations in the Enhanced Vegetation Index (3rd row and the corresponding aerial view; 4th row) to quantify vegetation properties.

Specifically, individual flux tower observations represent an aggregate of surface conditions over a dynamically varying upwind area (termed flux footprint). Yet, the traditional approach to constrain ecosystem carbon budgets simply aggregates observations over certain periods (Aubinet, Vesala, and Papale 2012; Lee, Massman, and Law 2004), regardless of flux footprint variations. In particular, over heterogeneous ecosystems (Fig. 1), this systematically embeds variability that has not been previously accounted for (Amiro 1998; Griebel et al. 2016; Leclerc, Meskhidze, and Finn 2003; Nappo et al. 1982; Schmid et al. 1999). This calls into question the accuracy of upscaling methods, as the contributions of such surface heterogeneities at the model grid scale are rarely quantified or attributed (Chen et al. 2009; Rebmann et al. 2005).

Unfortunately, there is no standard method currently in place to rectify the magnitude of this interactive effect. In this paper we (i) assess the prevalence and magnitude of the spatial heterogeneity of vegetation characteristics within a 500 m radius of all Tier 1 sites from the FluxNet2015 data set; (ii) assess the sampling coverage of fluxes in space and time; (iii) provide a practical solution that treats the embedded spatial sampling bias by attributing the contributions of spatial heterogeneities to flux tower estimates. This allows us to (iv) assess the magnitude of the effects of spatial sampling corrections on mean biome productivity estimates.

The proposed approaches to quantify surface properties and spatio-temporal sampling characteristics provide the first standardized methods of assessing and reporting site homogeneity within the flux community (Fig. 2). By removing this unaccounted variability through variations in spatial and temporal sampling patterns over heterogeneous sites we can attribute the contributions of surface heterogeneities to flux tower estimates. This enables us to establish more reliable impacts of changing weather patterns on ecosystem functioning and which will ultimately result in more accurate predictions of global productivity in future climates.

2 Materials and methods

2.1 Heterogeneity Index

We constrained a Heterogeneity Index (HI) based on the EVI (Enhanced Vegetation Index) within a 500 m radius around each flux tower from Landsat 5 (1984-2011) and Landsat 8 (2014) using Google Earth Engine. The 500 m radius (r) was an example, an illustrative compromise between sites with a potentially smaller footprint (e.g. croplands and grasslands) and sites with a potentially larger footprint (tall towers and forest sites), as the lack of site-specific information about canopy and instrument height (and how these varied through time) prevented us from calculating actual source areas of fluxes for each site year. Landsat 7 data was discarded due to the Scan Line Corrector failure and only years with complete coverage were considered. All scenes were masked for clouds, cloud shadows and snow presence. Next, for all years where Landsat data was available, we calculated annual summary statistics (median, percentiles) for each 30x30 m pixel within the 500 m radius around the flux tower location. The HI of each year was then calculated from the difference between the 95th and 5th percentile divided by the median of pixels in the 500 m radius:

$$HI(r) = \frac{EVI_{P95}(r) - EVI_{P5}(r)}{EVI_{P50}(r)} \quad (1)$$

Lastly, we calculated the mean site HI as the arithmetic mean of all years with flux observations in the FLUXNET2015 data base.

2.2 Eddy covariance data

We used all 154 Tier 1 sites that supplied wind characteristics with their flux data from the FluxNet2015 data set to calculate the spatio-temporal sampling coverage (section 2.2.1). We excluded biomes that were only represented by one site in the database (closed shrublands; CSH). To ensure that disproportionally large data gaps in a given year would not bias the observations towards specific seasons, we removed years with less than 50% of the site-specific mean annual data coverage (72 site years from 64 sites), as these also resulted in large deviations from the mean site and mean biome NEP (Figs. S1 and S2). From the remaining 1071 site years we calculated the mean traditional and space-time-equitable carbon budgets per site (section 2.2.3), and grouped the 154 sites into 10 biomes following the IGBP classification that was provided with the data. It should be mentioned that our data filtering yielded more conservative results, i.e. smaller differences between traditional and space-time-equitable budgets. Further, it should be noted that we are not attempting to generate a true site budget when using only observational data, as these are biased towards daytime observations (see section 4.3.1). We only provide estimates from non-gap-filled data to get a number that allows us to assess the magnitude of impact of the way we constrain carbon budgets (that is by comparing budgets based on simple sums of observations with the space-time-equitable approach that treats the spatio-temporal sampling bias), as these are based on exactly the same underlying data for all methods (Griebel et al. 2016; Horst et al. 2019). We provide an evolution of our accounting approaches (SI section 2) in the Supplementary Information.

2.2.1 Spatio-temporal sampling coverage

The annual spatio-temporal sampling coverage (SC) is calculated from all direct observations in a given year (Fig. S3 and SI section 1) by:

1. attributing the relative contribution of observations in a matrix with wind sectors as columns (spatial bins), and time intervals as rows (temporal bins; Fig. S3a,b);

2. summing across each column/row to isolate the spatial from the temporal sampling variability (Fig. S3c,d);

3. arranging row/column means (in ascending order) and plotting the cumulative sums of observed against evenly distributed relative contributions (Fig. S3e,f).

4. Individual spatial/temporal SC is then defined as the divergence from the 1:1 line (ratio of areas under curves; Fig. S3e,f); and

5. the composite spatio-temporal SC is derived as the mean from the temporal and spatial components.

Lastly, we calculated the mean spatio-temporal SC for each site as the arithmetic mean of the all years with flux observations in the FLUXNET2015 data base. The SC ranges from 0% (unilateral sampling in space and time, i.e. all observations derived from one wind sector and one time interval) to 100% (i.e. fully balanced sampling in space and time with equal contribution from each wind sector and time interval; Fig. S3e,f). Note that arranging row/column means in ascending order puts more weight on bins with low coverage. However, this is necessary to remove the variation of sampling coverage due to the location of the sectors with low coverage (Fig. S4). While this puts more weight on bins with lower than expected coverage and creates a worst-case scenario for the sampling coverage, an ascending arrangement presents an objective and practical indication that a sampling bias exists when calculating carbon budgets at heterogeneous sites with a low temporal or spatial sampling coverage (Fig. S5).

2.2.2 Traditional carbon budgets

Traditional budgets assume homogeneous surface characteristics (Aubinet et al. 2012; Lee et al. 2004) and observations are aggregated over certain periods regardless of flux footprint variations due to changes in prevailing winds (Fig. 1). For each FluxNet site we used all available observational data that passed the ustar-filter (NEE_VUT_REF conditional to that NEE_VUT_REF_QC=0), as well as gap-filled data (NEE_VUT_REF conditional to NEE_VUT_REF_QC=3) to calculate traditional annual budgets (Φ) by aggregation of each flux observation (F_c) in a given year (i):

$$\Phi(i) = \sum_{i=1}^{k} F_c(i) \tag{2}$$

2.2.3 Space-Time-equitable carbon budgets

Building on the original approach that attributes spatial sampling variations to carbon budgets based on average wind patterns by Griebel et al. 2016 (SI section 2.1.1 and Figs. S6, S7), we developed space-time-equitable carbon budgets to equally attribute spatial and temporal sampling variations from an uneven spatio-temporal sampling coverage (SI section 2.1.3). Space-time-equitable budgets attribute equally in space (i.e., each wind sector contributes an equal amount to the annual budget; see also space-equitable budgets in SI section 2.1.2) as well as in time (i.e., the mean flux in each wind sector was comprised of equal time slots to ensure an equal diurnal representation in each wind sector). Space-time-equitable budgets are calculated in three steps:

1. attributing the mean carbon flux ($\overline{F_c}(i,j,h)$) in a matrix with wind sectors (j) as columns (k spatial bins), and time intervals (h) as rows (m temporal bins) during each year (i)

2. calculating the mean space-time-equitable carbon flux (ϕ) by calculating row means followed by the column mean of the matrix:

$$\phi(i) = \sum_{h=1}^{m} \sum_{j=1}^{k} \overline{F_c}(i,j,h) * \frac{1}{k\ m} \tag{3}$$

3. multiplying the mean carbon flux in each year with the total number of observations in that year:

$$\Phi(i) = \Omega(i) * n(i) \tag{4}$$

This ensures comparability between traditional budgets that were based on the sum of all observations and space-time-equitable budgets.

2.3 Statistical approach

Significant relationships between the mean annual HI and the absolute difference between traditional and space-time-adjusted budgets were tested using a linear model for all site-years and site means that passed the quality control, using observational as well as gap-filled data (Figs. S8-S10). We performed a two-way ANOVA and a Tukey test to assess significant differences in estimates of mean biome NEP between traditional and space-time-equitable budgets, as well as their interaction. The underlying assumptions were assessed using qqplot and residual plots. Data for traditional budgets were log-transformed and space-time-equitable budgets were square-root transformed to correct for non-normal data distributions. Residual plots confirmed constant variance.

3 Results

3.1 Quantifying spatial heterogeneity of vegetation and sampling coverage around flux towers

We identified variability of vegetation characteristics at the majority of flux sites, as indicated by a mean Heterogeneity Index (HI) of 0.58 (P_{25}=0.29 and P_{75}=0.67; Fig. 2a). Variations in vegetation characteristics were present in all biomes (Fig. 2a), with a larger HI at sites in Europe and at high latitudes compared to sites in the central US and the southern hemisphere. Evergreen broadleaf forests and Savanna's tended to be more homogeneous, whereas wetlands, grasslands, open shrublands and mixed forests were characterized by a higher HI. Further, a mean spatio-temporal sampling coverage (SC) of

70.6% indicated variations in the uniformity of the spatial as well as the diurnal sampling pattern at most sites (P_{25}=65.5% and P_{75}=76.8%; Fig. 2b), which have the potential to introduce variations in annual budget estimates depending on the parts of the ecosystem being sampled at heterogeneous sites. To date, flux footprint variations over heterogeneous ecosystems due to changes in prevailing winds (Fig. 1) have not been accounted for when generating annual budgets with the only existing and universally applied traditional budgeting approach. Thus, we propose a practical solution to reduce the embedded spatial and temporal sampling biases by attributing the contributions of spatial heterogeneities to flux tower budgets (Figs. S6, S7).

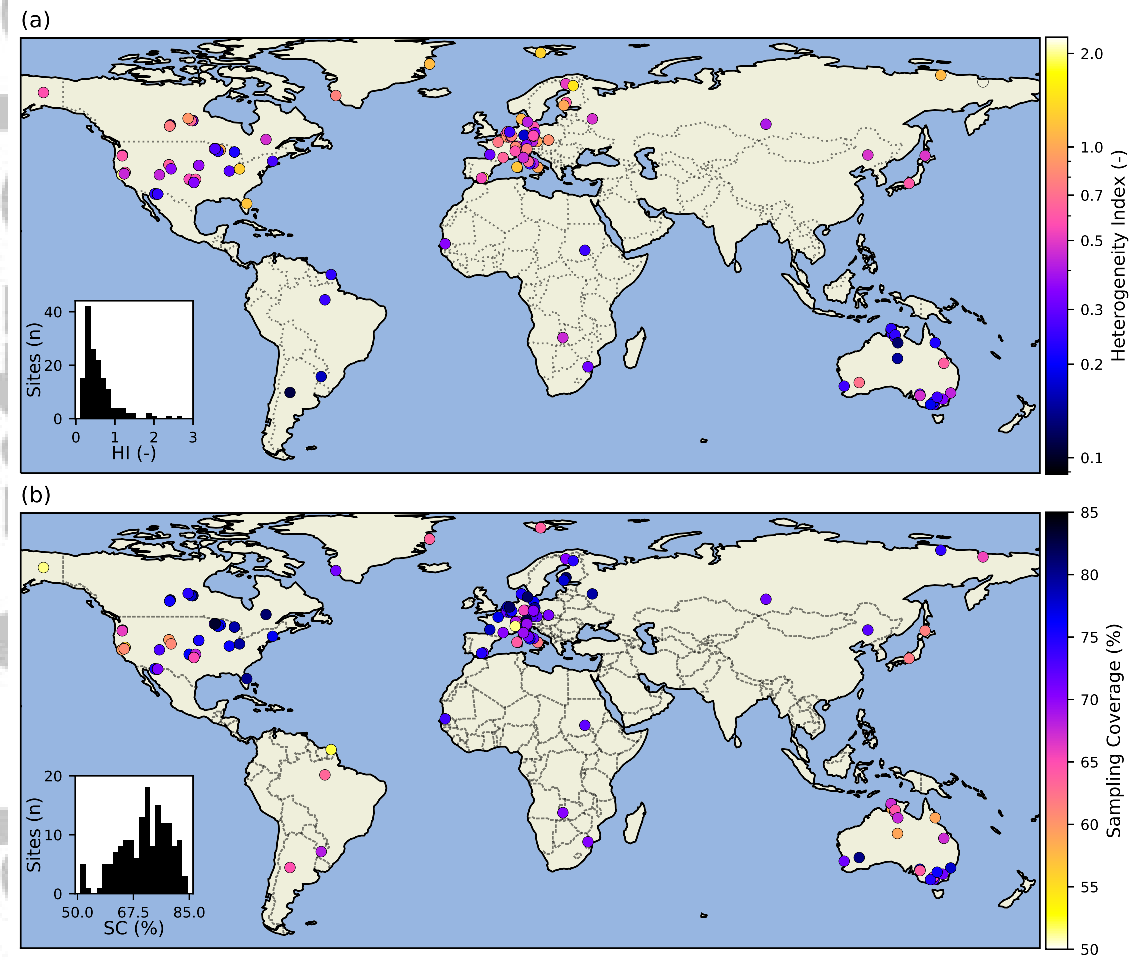

Figure 2: Spatial distribution of Tier 1 sites that contributed data to the FluxNet2015 dataset (154 sites, Table S1). Color represents a) the site heterogeneity as indicated by the mean annual Heterogeneity Index (HI), and b) the spatio-temporal sampling coverage (SC) across all observation years that were contributed to the FluxNet dataset. Larger values indicate larger site heterogeneity or more uniform sampling coverage.

3.2 The way we budget matters for biome productivity estimates

Across all sites, the relationship between the mean annual HI and the absolute difference between traditional and space-time-equitable budgets was significant, even when using gap-filled data($P<0.05$; Fig. S8). While the relationship for observational data is still affected by the imbalanced diurnal sampling pattern Chu, Baldocchi, John, Wolf, and Reichstein

2017 and confounded by generally small budgets in the identified heterogeneous biomes; Figs. S8, S9), a positive relationship persists when using gap-filled data (which despite not accounting for spatial sampling variations does remove the diurnal sampling bias; Figs. S8, S10). These non-negligible differences between traditional and space-time-equitable budgets demonstrate that the size of fluxes was not uniform with wind direction at the majority of sites (Fig. 2), and highlights that a significant share of a sites annual budget is determined by how often each part of the ecosystem has been sampled in a given year (Figs. 1, 2). These site-level effects are then transferred to mean biome NEP estimates, which differ significantly between traditional and space-time-adjusted budgets ($P<0.001$ using observed and $P<0.05$ using gap-filled data; Fig. 3).

Even for traditionally constrained carbon budgets, there is considerable variation in traditionally calculated mean annual NEP among sites in each biome represented in the FluxNet2015 data set, irrespective of whether space-time-equitable NEP was based on observational ($P<0.001$; Fig. 3) or on gap-filled data ($P<0.05$; Fig. S11). On average, across all FluxNet2015 sites, mean biome NEP for space-time-equitable budgets differed from traditional budgets by -10.4% using gap-filled data (ranging from + 12.6% at Savanna's to -35.1% at wetlands; Table S2), highlighting the non-negligible variation of NEP when constraining budgets as simple sums. This difference is evident despite a low mean data coverage (34% of high quality data per site; Fig. S12) and despite gap-filling procedures assuming homogeneous site characteristics (they have 100% temporal sampling coverage, but are not stratified by wind direction and thus do not consider spatial sampling variations over heterogeneous ecosystems). The difference between the traditional and the space-time-equitable budgeting approach to derive the mean annual biome NEP increases significantly when using only observational data (-49.6%, ranging from -36% at mixed forests to -62% at grasslands; Fig. 3, Table S2), which results from the disproportional diurnal data coverage at individual sites and the proportional re-weighting of nighttime data in the space-time-equitable budgets that introduces variability in the temporal sampling coverage (Fig. S3).

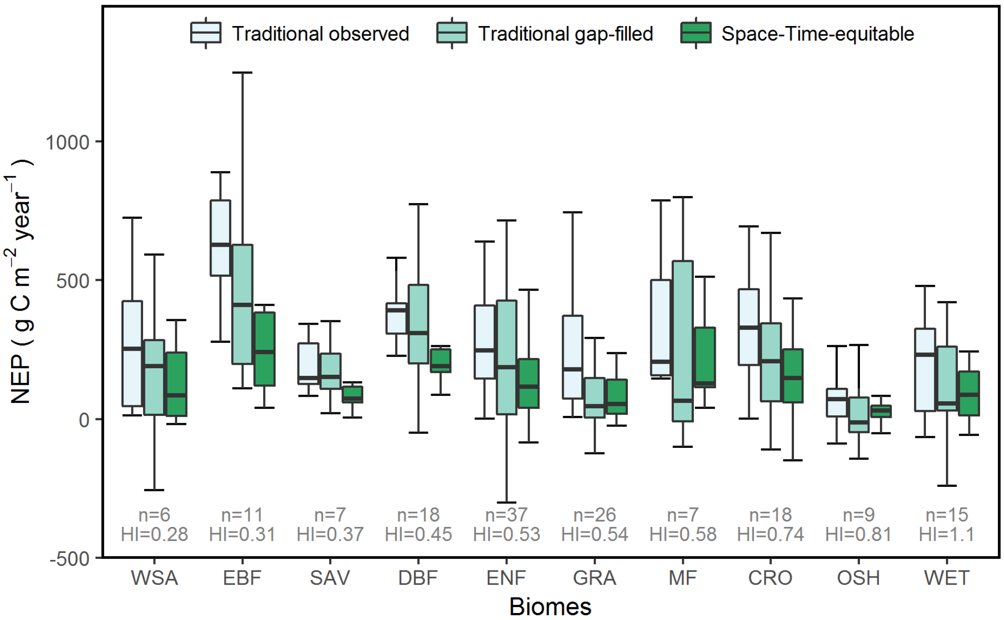

Figure 3: Variability of mean annual net ecosystem productivity (NEP; gCm^{-2}yr^{-1}) estimates depending on the method that was used to constrain the mean carbon sequestration rates for each site in each biome. Traditional budgets represent the unweighted sum of all observations in a year using only directly measured (traditional observed) and gap-filled data (traditional gap-filled). Space-time-equitable budgets (using observational data only) weighted all direct observations towards a spatially equal contribution of each wind sector and an equitable temporal representation. Note that the data used to constrain the mean annual NEP for each site included all site-years that passed the minimum data coverage requirement, and that individual years were averaged to a site mean prior to grouping by biome. Whisker length indicates the inter-site variability as the upper and lower range of the 95% confidence interval, the line represents the median annual NEP for each biome, and the biome order reflects the mean biome Heterogeneity Index (HI; increasing towards the right). Numbers indicate the total number of sites in each biome: WSA=woody savannah, EBF=evergreen broadleaf forest, SAV=savannah, DBF=deciduous broadleaf forest, ENF=evergreen needleleaf forest, GRA=grassland, MF=mixed forest, OSH=open shrublands, WET=wetland.

4 Discussion

4.1 A practical solution to treat the embedded spatial sampling bias

Traditional annual carbon budgets comprise a simple sum of gap-filled tower observations over a year, assuming perfectly homogeneous conditions in space and time (Aubinet et al. 2012; Lee et al. 2004). They are typically based on gap-filled data, yet existing gap-filling procedures assume homogeneous site characteristics as they are not stratified by wind direction, and thus do not consider flux footprint variations over heterogeneous ecosystems. Further, with an average data coverage of 34% for high quality data (Fig. S12), the number of modeled data far outweighs the number of observations in a given year at most sites, which is particularly prevalent in very cold environments (Horst et al. 2019). However, all sites to some degree (Aubinet et al. 2012; Baldocchi 2003), and some sites to a very large degree, experience diurnal variations in the prevailing wind directions, e.g. sites affected by land-sea breezes or by mountain-valley winds (Goulden, Munger, Fan, Daube, and Wofsy 1996; Hiller, Zeeman, and Eugster 2008; Zardi and Whiteman 2013).

Earth system models typically operate at a larger scale than flux tower footprints (∼3000 km^2 vs. 1-10 km^2; Jung, Reichstein, and Bondeau 2009), but they assume that flux tower observations are representative of the mean biophysical and climatological properties of the grid cells they are located in. Space-time-equitable budgets provide a more uniform spatial average of the ecosystems flux characteristics than traditional budgets, through rectifying variations in the sampled upwind area. Furthermore, the equal attribution of temporal bins within each sector ensures that the mean flux of each wind sector is based on a time-uniform contribution of data (Fig. S6). Thus, space-time-equitable budgets are expected to improve the spatial representativeness of an ecosystem's carbon budget, as each wind sector contributes the same proportion to an annual budget irrespective of how often it has been sampled in any given year (Fig. S6). Such spatially robust productivity estimates allow for more accurate examination of interannual variations and long-term trends of flux tower data, while also being more appropriate for linking fluxes with gridded climate and remote sensing data.

4.2 Systematic attribution of surface heterogeneities to carbon budgets

We identified variability of vegetation characteristics at the majority of flux sites (mean HI=0.58; Fig. 2), providing evidence that spatial homogeneity of site characteristics does not hold true for many flux tower locations. A perfectly homogeneous ecosystem would have a HI of zero. Likewise, a synchronous change of all pixels throughout the ecosystem over a year (e.g. deciduous forests) results in similar mean annual values of the individual pixels and ultimately an HI of zero. The index only increases with increasing difference between the upper and lower percentiles of pixel values in respect to the median value across all pixels. That is, in a heterogeneous ecosystem (e.g. mixed forests), individual pixel values might change at different rates throughout the year, resulting in spatial gradients across individual mean pixels and a non-zero HI. The HI might be extended by incorporating variations in topography and soil properties, and ideally each site would use a footprint model to constrain the exact spatial extent. Yet, our HI provides the first standardized measure for the flux community to report site heterogeneity along with carbon budgets. Further, the HI might not be zero, but the spatio-temporal sampling coverage might be uniform (100%). Then, the balanced sampling pattern would fully account for the heterogeneous surface characteristics, and space-time equitable budgets would be identical to traditional budgets. However, no site was fully homogeneous and no site had a mean spatio-temporal sampling coverage above 85% (Fig. 2), suggesting that carbon budget estimates are commonly affected by sampling variations at most sites to a small degree, and at some sites to a large degree.

Spatial heterogeneity and its effect on CO_2 budgets have been the subject of very few targeted studies (Griebel et al. 2016; Leclerc et al. 2003; Nappo et al. 1982; Schmid et al. 1999). To our knowledge, ours is the first study to quantify the effect of spatio-temporal sampling variations over heterogeneous ecosystems on annual CO_2 sequestration estimates at flux tower sites. Further, our findings indicate that heterogeneity-induced uncertainty from individual tower observations affects mean biome productivity estimates on orders up to 35% (gap-filled data; Table S2) and 62% (observational data; Fig. 3, Table S2). Thus, if left untreated, even sophisticated upscaling of flux tower observations via remote sensing and Earth system modelling are degraded through superimposing spatio-temporal sampling patterns over actual biophysical variability at the measurement sites, thus contributing to uncertainties in global carbon budget estimates (Ahlström et al. 2015; Moncrieff, Malhi, and Leuning 1996; Schimel et al. 2001). As a solution we suggested alternative constraints for calculating ecosystem carbon budgets by systematically attributing the contributions of surface heterogeneities and variations in the spatio-temporal sampling coverage to annual carbon budgets.

4.3 Limitations and trade-offs

4.3.1 Using observational vs. gap-filled data

Over the past decades, significant progress has been made by the flux community to establish rigorous quality control assessments that test for the fulfillment of theoretical requirements and sensor and deployment related issues (Foken and Wichura 1996; Lee et al. 2004), providing a variety of tools to flag, correct and gap-fill affected data (Metzger et al. 2017; Wutzler et al. 2018). Keeping data comparable for synthesizing observations between sites requires stringent quality control, which in the case of the FluxNet2015 data set retained on average 34% of observations in a yearly file (1300 site years; Fig. S12). This, by implication, means that the majority of flux data at most sites has been gap-filled, which obscures the very effect we aimed to investigate in this study (Fig. S11). While focusing on observational data precluded us from generating a true site budget due to a skewed temporal coverage towards daytime data at most sites (Fig. S3), it still allowed us an unbiased assessment on the magnitude of the impact of the way we constrain a site's annual budget, since the resulting traditional and space-time-equitable budgets were based on exactly the same underlying data.

Presenting the sampling coverage and carbon budgets using a full year of data was a trade-off to ensure sufficient data coverage in each bin. Though monthly aggregation periods might be robust to small data gaps and more suitable for sophisticated upscaling and benchmarking of Earth system models (the difference between monthly and annual space-time-equitable budgets was insignificant for observational and gap-filled data; Fig. S13), they remain problematic with longer data gaps in excess of a month, which is not uncommon for remote sites and unavoidable following instrument failures. Still, this trade-off resulted in uneven data coverage in spatial and temporal bins for space-time-equitable budgets, and ultimately in non-uniform uncertainties associated with each bin that depend on the individual spatio-temporal sampling coverage at each site. Further, we need to acknowledge that our treatment of biases resulting from upscaling heterogeneity-induced errors follows a simple tile-approach (Mengelkamp et al. 2006), by considering spatio-temporal proportions through aggregation. In remote sensing and Earth system modeling applications, more sophisticated approaches to utilize flux observations include e.g. ensemble Kalman filtering (Fox et al. 2009). These approaches establish a precision-criterion for individual flux observations that can help attenuate some of the effects reported here. Yet, unless assimilation is performed in a fully spatio-temporally explicit frame of reference

(Metzger 2018), operational approaches are still subject to similar heterogeneity-induced biases as reported here.

4.3.2 Overcoming the scale-mismatch between the flux and remote sensing domains

While our suggested approaches improve the accuracy of upscaling flux tower carbon budgets to model grid cells, we are still unable to overcome the scale-mismatch between flux tower observations and Earth system models grid cells (Jung et al. 2009). For instance, the tower may not be centered in the respective grid cell used for scaling, or the tower footprint may have been intentionally constrained to only sample a certain land cover type and thus does not sample all portions of a heterogeneous grid cell. Development of general solutions to transform traditional flux observations to spatially representative estimates at modeling grid scale is ongoing (e.g. Metzger 2018; Xu, Metzger, and Desai 2018) and often require sophisticated data sources and processing. Further, space-time-equitable budgets remain spatially constrained by the footprint extent of the tower, that is, they still are unable to provide a true grid-scale average as is used for remote sensing and Earth system modeling. Nonetheless, they provide a more realistic picture of the average grid cell properties surrounding the flux tower. Furthermore, using only observational data indicates a consistent overestimation of ecosystem productivity due to the combination of equal diurnal contributions with the attribution of heterogeneous site characteristics. However, the unilateral direction of consistent overestimation of mean biome NEP became bi-directional when using gap-filled data (which has a uniform temporal sampling coverage), thus leaving an equal spatial attribution as the only difference to traditional budgets. The remaining difference of up to 62% of mean annual NEP estimates between space-time-equivalent and traditional budgets highlights the importance of quantifying and attributing the spatial heterogeneity of surface properties for flux tower budgets, and promises substantial increases in the accuracy of terrestrial productivity estimates when incorporating spatio-temporal carbon budgets into sophisticated gap-filling and scaling approaches.

In this study we made a first attempt to propose a standardized index to assess and report the heterogeneity of vegetation characteristics of flux sites using remotely sensed satellite images, and to quantify the spatio-temporal sampling variation using readily available flux tower data. Our findings demonstrate that estimated NEP of flux tower sites can be severely compromised when using traditional budgeting and data assimilation approaches, and highlights the urgent need to update the network-wide standard procedures for gap-filling missing data and for computing carbon budgets. While our study focused on providing evidence that surface heterogeneities significantly affect carbon budgets at the site and the biome scale, follow-up assessments are required that thoroughly disentangle the effects and uncertainties associated with the spatial and temporal binning size on adjusted carbon budgets. However, our approach to account for spatial heterogeneities when computing carbon budgets bridges a critical gap in scaling ecosystem processes from individual sites to the globe that holds considerable promise in allowing for a more accurate attribution of climate change mitigation by terrestrial carbon uptake.

Supporting Information (SI)

Detailed methods, including calculation of the spatio-temporal sampling coverage (section 1), the evolution of scaling approaches (section 2) and Figs. S1-S13, Tables S1-S2.

Acknowledgements

We thank two reviewers for their valuable feedback. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices. Eddy covariance data is available under http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/. This research was supported by Western Sydney University and TERN-Ecosystem Processes, via the National Collaborative Research Infrastructure Strategy Program (NCRIS). The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by Battelle. This material is based in part upon work supported by the National Science Foundation [grant DBI-0752017]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors declare no competing financial interests.

Author contributions

AG and DM designed the synthesis study with input from GB, SM and EP. AG and DM analyzed the data and calculated the carbon budgets and the spatio-temporal sampling coverage, DM implemented the analysis of the Heterogeneity Index in GoogleEarthEngine.

AG drafted the manuscript and all authors contributed to writing the final version.

Data availability

Eddy covariance data is available under http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/. See Supporting Information for equations and instructions on how to calculate carbon budgets, heterogeneity index and spatio-temporal sampling coverage. The source code to calculate space-time-equitable carbon budgets is available from GitHub (https://github.com/AnneGriebel/Griebel-GRL_2020), including the eddy4R GitHub repository (https://github.com/NEONScience/eddy4R-documentation/wiki/What-is-eddy4R) and the eddy4R Docker repository (https://hub.docker.com/r/stefanmet/eddy4r/).

References

Ahlström, A. et al. (2015). The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. *Science* 348 (6237), 895–899.

Amiro, B. (1998). Footprint climatologies for evapotranspiration in a boreal catchment. *Agricultural and Forest Meteorology* 90 (3), 195–201.

Aubinet, M., T. Vesala, and D. Papale (2012). Eddy covariance: a practical guide to measurement and data analysis. Springer Science & Business Media.

Baldocchi, D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future. *Global Change Biology* 9 (4), 479–492.

Baldocchi, D. (2008). Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany* 56 (1), 1–26.

Bodesheim, P., M. Jung, F. Gans, M. Mahecha, and M. Reichstein (2018). Upscaled diurnal cycles of landatmosphere fluxes: a new global half-hourly data product. *Earth System Science Data* 10 (3), 1327–1365.

Chen, B., T. Black, N. Coops, T. Hilker, J. Trofymow, and K. Morgenstern (2009). Assessing Tower Flux Footprint Climatology and Scaling Between Remotely Sensed and Eddy Covariance Measurements. *Boundary-Layer Meteorology* 130 (2), 137–167.

Chen, B. et al. (2012). Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis. *Remote Sensing of Environment* 124, 742–755.

Chen, M. et al. (2017). Regional contribution to variability and trends of global gross primary productivity. *Environmental Research Letters* 12 (10), 105005.

Chu, H., D. D. Baldocchi, R. John, S. Wolf, and M. Reichstein (2017). Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. *Journal of Geophysical Research: Biogeosciences* 122 (2), 289–307.

Foken, T. and B. Wichura (1996). Tools for quality assessment of surface-based flux measurements. *Agricultural and Forest Meteorology* 78 (1-2), 83–105.

Foken, T. and M. Leclerc (2004). Methods and limitations in validation of footprint models. *Agricultural and Forest Meteorology* 127 (3-4), 223–234.

Fox, A. et al. (2009). The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology* 149 (10), 1597–1615.

Giannico, V., J. Chen, C. Shao, Z. Ouyang, R. John, and R. Lafortezza (2018). Contributions of landscape heterogeneity within the footprint of eddy-covariance towers to flux measurements. *Agricultural and Forest Meteorology* 260-261, 144–153.

Goeckede, M. et al. (2008). Quality control of CarboEurope flux data - Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems. *Biogeosciences* 5 (2), 433–450.

Goulden, M. L., J. W. Munger, S. M. Fan, B. C. Daube, and S. C. Wofsy (1996). Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global Change Biology* 2 (3), 169–182.

Griebel, A., L. T. Bennett, D. Metzen, J. Cleverly, G. Burba, and S. K. Arndt (2016). Effects of inhomogeneities within the flux footprint on the interpretation of seasonal, annual, and interannual ecosystem carbon exchange. *Agricultural and Forest Meteorology* 221, 50–60.

Hiller, R., M. J. Zeeman, and W. Eugster (2008). Eddy-covariance flux measurements in the complex terrain of an Alpine valley in Switzerland. *Boundary-Layer Meteorology* 127 (3), 449–467.

Horst, S. V. van der, A. J. Pitman, M. G. D. Kauwe, A. Ukkola, G. Abramowitz, and P. Isaac (2019). How representative are FLUXNET measurements of surface fluxes during temperature extremes? *Biogeosciences* 16 (8), 1829–1844.

Huang, K. et al. (2018). Enhanced peak growth of global vegetation and its key mechanisms. *Nature Ecology & Evolution* 2 (12), 18971905.

Jung, M., M. Reichstein, and A. Bondeau (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* 6 (10), 2001–2013.

Jung, M. et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research* 116.

Kim, J., Q Guo, D. Baldocchi, M. Leclerc, L Xu, and H. Schmid (2006). Upscaling fluxes from tower to landscape: Overlaying flux footprints on high-resolution (IKONOS) images of vegetation cover. *Agricultural and Forest Meteorology* 136 (3-4), 132–146.

Leclerc, M., N. Meskhidze, and D. Finn (2003). Comparison between measured tracer fluxes and footprint model predictions over a homogeneous canopy of intermediate roughness. *Agricultural and Forest Meteorology* 117 (3-4), 145–158.

Lee, X., W. Massman, and B. Law (2004). Handbook of Micrometeorology: a guide for surface flux measurement and analysis. Vol. 29. Springer Science & Business Media.

Mengelkamp, H.-T. et al. (2006). Evaporation over a heterogeneous land surface. *Bulletin of the American Meteorological Society* 87 (6), 775–786.

Metzger, S. (2018). Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations. *Agricultural and Forest Meteorology* 255, 68–80.

Metzger, S. et al. (2017). eddy4R 0.2. 0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5. *Geoscientific Model Development* 10, 3189–3206.

Moncrieff, J., Y. Malhi, and R. Leuning (1996). The propagation of errors in longterm measurements of landatmosphere fluxes of carbon and water. *Global Change Biology* 2 (3), 231–240.

Nappo, C. et al. (1982). The Workshop on the Representativeness of Meteorological Observations, June 1981, Boulder, Colo. *Bulletin of the American Meteorological Society* 63 (7), 761–764.

Novick, K. et al. (2018). The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology* 249, 444–456.

Rebmann, C. et al. (2005). Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling. *Theoretical and Applied Climatology* 80 (2-4), 121–141.

Schimel, D. S. et al. (2001). Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* 414 (6860), 169.

Schmid, H. P. and C. R. Lloyd (1999). Spatial representativeness and the location bias of flux footprints over inhomogeneous areas. *Agricultural and Forest Meteorology* 93 (3), 195–209.

Stoy, P. C. et al. (2013). A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and Forest Meteorology* 171, 137–152.

Wutzler, T. et al. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences* 15 (16), 5015–5030.

Xu, K., S. Metzger, and A. R. Desai (2018). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. *Agricultural and Forest Meteorology* 255, 81–91.

Zardi, D. and C. D. Whiteman (2013). Diurnal mountain wind systems. *Mountain Weather Research and Forecasting*. Springer, p. 35–119.