

Review of the manuscript essd-2016-36

Title: “Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements”

by Kumar et al.

General Comments

The main question addressed by the research is how well global terrestrial carbon cycling with the atmosphere can be represented over time when informed by time-resolved FLUXNET eddy-covariance observations and static delineations of eco-climatic regions. Corresponding goals are to understand the spatio-temporal representativeness of FLUXNET, to explore the patterning of inter-model agreement, and to identify eco-climatic regions that warrant future focus. The global mapping of FLUXNET carbon cycle observations is achieved through inverse distance weighted interpolation: The Euclidian distance between a target pixel and suitable FLUXNET sites is determined in multi-dimensional climate state-space, and not as a function as geographic proximity. One main difference to similar studies mapping the global carbon cycle is that no additional time-resolved drivers such as satellite data are used. The authors conclude that the study provides a window into the evolution of FLUXNET for representing the global carbon cycle, and that lower uncertainty is attained in the ecoregions that are well sampled.

The manuscript provides an incremental advance to the field of spatial representativeness and carbon cycle mapping. While the central questions are interesting and important, the manuscript essentially applies established methods to a new/different dataset. While the authors present conclusions about uncertainty metrics, these are not actually substantiated through quantification (details in the specific comments). To permit evaluating the significance and usefulness of the presented approach I strongly recommend a gridded uncertainty estimate from bootstrapping or propagation from cross-site validation. Special attention should be paid to the facts that (i) the Euclidean distance calculation does not appear to take into account inter-annual differences in environmental conditions, and (ii) FLUXNET sites are being switched on/off for interpolation based on monthly data availability. Hence, in the current form the results only partially contribute to and support the question and conclusions.

The manuscript clearly fits the aim of ESSD to further “the reuse of high-quality data of benefit to Earth system sciences”. However, in its current form two areas might exceed the journal scope, namely “any interpretation of data is outside the scope”, and “any comparison to other methods is beyond the scope”. Hence, I recommend to eliminate corresponding areas from the manuscript, and use the freed space for a thorough quantification and discussion of uncertainty. The manuscript length is appropriate in general, although Figure 1 and Figure 6 are partially redundant and can be combined. Writing and language is fluent, and no copy-editing is needed. However, it would be much desirable to provide sufficient information in the figure legends for interpretation without having to scour the entire text.

On this basis my recommendation is publishable in general, subject to major revisions.

Specific Comments

Please see in-text comments and edits below.



Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements

Jitendra Kumar¹, Forrest M. Hoffman², William W. Hargrove³, and Nathan Collier²

¹Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

²Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

³Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, Asheville, NC, USA

Correspondence to: Jitendra Kumar (jkumar@climatemodeling.org)

Abstract. Eddy covariance data from regional flux networks are direct in situ measurement of carbon, water, and energy fluxes and are of vital importance for understanding the spatio-temporal dynamics of the the global carbon cycle. FLUXNET links regional networks of eddy covariance sites across the globe to quantify the spatial and temporal variability of fluxes at regional to global scales and to detect emergent ecosystem properties. This study presents an assessment of the representativeness of FLUXNET based on the recently released FLUXNET2015 data set. We present a detailed high resolution analysis of the evolving representativeness of FLUXNET through time. Results provide quantitative insights into the extent that various biomes are sampled by the network of networks, the role of the spatial distribution of the sites on the network scale representativeness at any given time, and how that representativeness has changed through time due to changing operational status and data availability at sites in the network. To realize the full potential of FLUXNET observations for understanding emergent ecosystem properties at regional and global scales, we present an approach for upscaling eddy covariance measurements. Informed by the representativeness of observations at the flux sites in the network, the upscaled data reflects the spatio-temporal dynamics of the carbon cycle captured by the in situ measurements. This study presents a method for optimal use of the rich point measurements from FLUXNET to derive an understanding of upscaled carbon fluxes, which can be routinely updated as new data become available, and direct network expansion by identifying regions poorly sampled by the current network.

Data from this study are available at <http://dx.doi.org/10.15486/NGT/1279968>

1 Introduction

Terrestrial ecosystems are a critical component of the global carbon cycle regulating the climate through a range of biogeophysical and biogeochemical mechanisms. Terrestrial ecosystems exchange about 120 Pg of carbon per year with the atmosphere, through the processes of photosynthesis and respiration (Schlesinger and Bernhardt, 2013). Eddy covariance-based methods are widely used around the world to measure the gas exchange between vegetation and the atmosphere. The eddy covariance method assumes the measurement site is located in flat terrain, experiences steady or stable atmospheric conditions, and is surrounded by uniform vegetation for an extended distance in the upwind direction (Baldocchi, 2003). In practice, however, sites are often located in non-ideal and spatial heterogeneous regions and are thus may be prone to varying degrees of measurement

Commented [R1]: •It would be nice to include a few quantitative key results.



error (Baldocchi, 2003). Understanding the spatial representativeness of site measurements is required for landscape-scale integration and interpretation of carbon flux measurements. While an individual site may be considered representative of the biome in which it is located, any carbon flux measurement network must include a broad representation of vegetation types, climates, and stages of disturbance. FLUXNET is a global network of micrometeorological flux measurement networks consisting of individual sites that measure the exchange of carbon dioxide, water vapor, and energy between the biosphere and the atmosphere (Baldocchi et al., 2001). FLUXNET consists of more than ~750 sites (<http://fluxnet.fluxdata.org/sites/historical-site-status/>) across the globe that are independently operated by regional networks (<http://fluxnet.fluxdata.org/about/regional-networks/>). However, the locations of the sites in the network were not formally designed to uniformly and consistently observe global biomes and thus represent a sparse and spatially biased sampling of the global terrestrial ecosystem. Careful synthesis of spatiotemporally sparse flux observations is required to understand the regional to global scale carbon cycle. Upscaling of fluxes to regional to global scales requires understanding and quantification of the spatial representativeness of the global network of flux sites.

A number of studies have analyzed the representativeness of regional networks of flux sites. Hargrove et al. (2003) analyzed the representativeness of the AmeriFlux network using an ecoregion-based approach and found that while most of the continental United States was well represented by the network, the Pacific Northwest and Texas grasslands were poorly represented by the network sites. Yang et al. (2008) conducted a similar independent study of AmeriFlux network representativeness using remotely sensed climate and vegetation data products and arrived at conclusions similar to Hargrove et al. (2003). Hargrove and Hoffman (2005) presented a quantitative ecoregion-based approach to assess the representativeness of networks and a method for design of sampling networks. Sulkava et al. (2011) studied the representativeness of the European flux tower network and developed a quantitative network design tool for identifying and prioritizing addition of new flux sites. Chen et al. (2011, 2012) used remote sensing and footprint analysis to characterize the regional spatial representativeness of flux sites across the Canadian Carbon Program Network. He et al. (2015) conducted a similar assessment for the Chinese flux network. They also identified a set of sites that would complement the existing flux network. Most of these assessments of the spatial representativeness of regional flux networks across the globe have been focused on quantifying the coverage by the network and for the design of new sampling sites. However, the quantitative representativeness of network sites has not yet been applied to synthesize and upscale observations beyond the footprints of individual sites to provide landscape- to global-scale estimates of flux measurements.

We present here an analysis of the quantitative representativeness of the FLUXNET global network of sites that contributed to the FLUXNET2015 data set. While a large number of global flux sites are affiliated with the FLUXNET network, only a subset of site data are typically available in data collections. Thus, only the sites with data available in the most recent July 12, 2016, release of FLUXNET2015 were included in our study. Independently operated sites in the FLUXNET network often vary significantly in their period of active operation or data availability for logistical reasons. Thus in this study, in addition to quantifying the spatial representativeness of the network, we quantify the temporally variable representativeness of data availability within the FLUXNET2015 collection. Finally, we present an approach for global upscaling of eddy covariance fluxes from FLUXNET network informed by the temporally varying spatial representativeness of the flux sites.

Commented [R2]: •It would be good to clarify whether the incurred biases be taken into account in the presented approach, e.g. through filtering and footprint de-biasing.

Commented [R3]: •These are footprint-to-landscape assessments, and on a distinctly smaller scale compared to the other network-to-continent scale studies.
 •The first lends itself to also address forenamed EC methodological challenges “sites are often located in non-ideal and spatial heterogeneous regions and are thus may be prone to varying degrees of measurement error”
 •The latter mainly addresses the state-space coverage for continental-to-global scale flux budgeting exercises.
 •It would be good to provide a more clear distinction of these approaches in relation to the present manuscript. I.e., clarify whether EC methodological challenges will also be addressed.

Deleted: of regional representativeness

Commented [R4]: •Representativeness for what? Nappo et al. (1982) provides a good starting point for an unambiguous definition.
 •Nappo, C. J., Caneill, J. Y., Furman, R. W., Gifford, F. A., Kaimal, J. C., Kramer, M. L., Lockhart, T. J., Pendergast, M. M., Pielke, R. A., Randerson, D., Shreffler, J. H., and Wyngaard, J. C.: The workshop on the representativeness of meteorological observations, June 1981, Boulder, Colorado, Bull. Am. Meteorol. Soc., 63, 761-764, 1982.

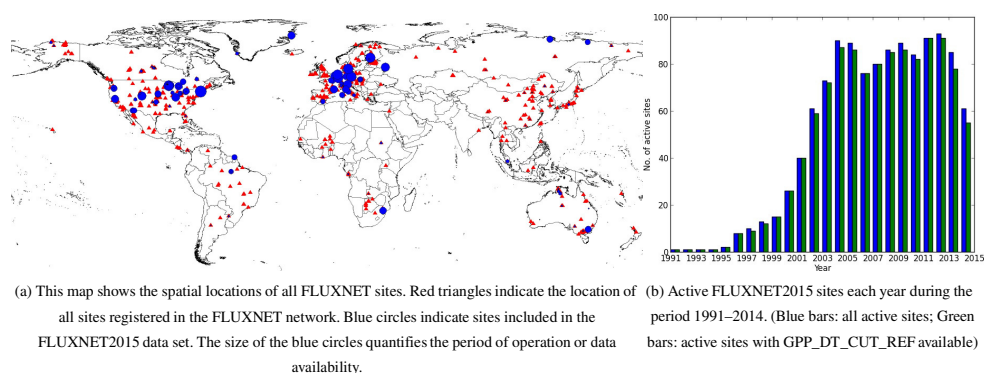


Figure 1. Eddy covariance sites in the FLUXNET2015 collection.

2 Data and Methods

2.1 Eddy covariance measurements

FLUXNET provides uniform and high quality data products through coordination among various regional flux networks across the globe (<http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/>). We used the latest release of the FLUXNET2015 eddy covariance data set collection in this study. The time series products in the FLUXNET2015 collection were gap filled (Reichstein et al., 2005) and went through a rigorous QA/QC procedure (Pastorello et al., 2014). The model efficiency method reference GPP product (GPP_DT_CUT_REF) calculated using day time data (Lasslop et al., 2010) was used to create an upscaled data product in this study.

While more than ~ 750 sites are registered in the FLUXNET database (<http://fluxnet.fluxdata.org/sites/historical-site-status/>), indicated by red triangles in Figure 1(a), data from only 165 global sites were available in the FLUXNET2015 data set released July 12, 2016, and shown by filled blue circles in Figure 1(b). This study was limited to data contained in the FLUXNET2015 data set. Environmental data sets used in our study were not available at the permanent wetland site at Adventdalen, Norway (NO-Adv) and thus was excluded from the analysis. Thus, the GPP observations from 164 sites globally (Figure 1(a)) spanning the period 1991–2014 were used. FLUXNET sites are maintained and operated by independent groups and the period of active operation and data availability varies widely from site to site, ranging from a single year to up to 24 years (Figure 1(a)). Figure 1(b) shows the number of active sites for which data was available in the FLUXNET2015 collection in any year during the period 1991–2014. Due to extremely limited data availability before year 2000, we limited our study to the period 2000–2014. There also exists some variability in variables and data fields available from any site. Figure 1(b) also shows the number of sites where the GPP_DT_CUT_REF variable used in this study was available.

Commented [R5]: •How are site-to-site differences and potential biases with regard to sensors, data acquisition and raw data processing being taken into account?
 •An uncertainty estimate from bootstrapping or propagation from cross-site validation into the global data product would be very helpful for judging the significance of results/differences.



2.2 Environmental variables

Ecosystem structure and functioning are governed by multiple independent control variables including climate, parent material, topography, potential biota and time (Chapin et al., 2012). A set of environmental variables to capture climatic, edaphic and topographic characteristics were selected to characterize the terrestrial ecosystem in this study (Table 1).

Table 1. Environmental variables used for ecoregion delineation, representativeness analysis and upscaling. These data are in the form of ~4 km raster grids.

Variable Description	Units	Source
Bioclimatic Variables		
Annual mean temperature	°C	Hijmans et al. (2005)
Mean diurnal range	°C	Hijmans et al. (2005)
Isothermality	–	Hijmans et al. (2005)
Temperature seasonality	°C	Hijmans et al. (2005)
Temperature annual range	°C	Hijmans et al. (2005)
Mean temperature of wettest quarter	°C	Hijmans et al. (2005)
Mean temperature of driest quarter	°C	Hijmans et al. (2005)
Mean temperature of warmest quarter	°C	Hijmans et al. (2005)
Mean temperature of coldest quarter	°C	Hijmans et al. (2005)
Annual precipitation	mm	Hijmans et al. (2005)
Precipitation during the wettest quarter	mm	Hijmans et al. (2005)
Precipitation during the driest quarter	mm	Hijmans et al. (2005)
Precipitation during the warmest quarter	mm	Hijmans et al. (2005)
Precipitation during the coldest quarter	mm	Hijmans et al. (2005)
Edaphic Variables		
Available water holding capacity of soil	mm	Global Soil Data Task Group (2000); Saxon et al. (2005)
Bulk density of soil	g/cm ³	Global Soil Data Task Group (2000); Saxon et al. (2005)
Soil carbon density	g/m ²	Global Soil Data Task Group (2000); Saxon et al. (2005)
Total nitrogen density	g/m ²	Global Soil Data Task Group (2000); Saxon et al. (2005)
Topographic Variables		
Compound topographic index (relative wetness)	–	Saxon et al. (2005)

Hijmans et al. (2005) developed very high resolution interpolated climate surfaces (WorldClim) for global land areas. Using a large number of global sources of weather station data, they created global climate surfaces at 30 arc second spatial resolution (~1 km). They also derived a set of 19 bioclimatic variables (BioClim) (with spatial resolution of 2 arc minute or ~4 km) that are biologically meaningful, representing annual trends, seasonality and extreme or limiting environmental factors. We selected 14 bioclimatic variables (Table 1) to represent the global bioclimatic environment in this study. We also selected four edaphic variables (Global Soil Data Task Group, 2000; Saxon et al., 2005) and one topographic variable (Saxon et al., 2005) to characterize not just the above-ground but also the below-ground environment.

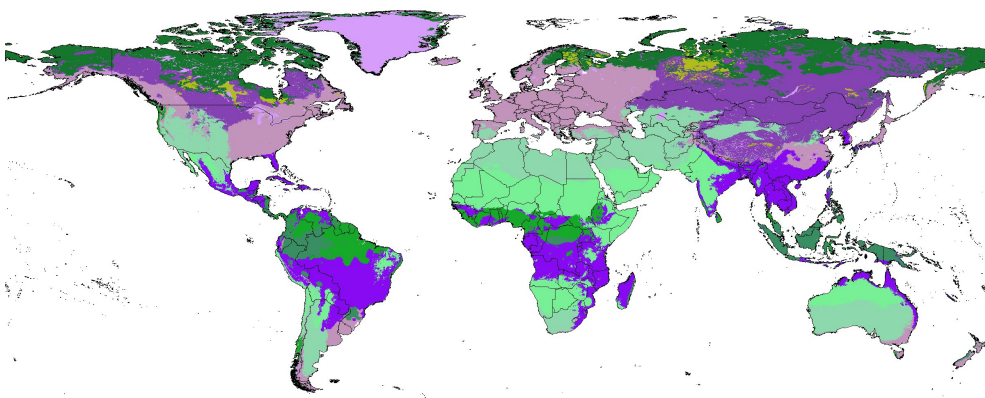


Figure 2. Global ecoregions delineated using 20 bioclimatic, edaphic and topographic characteristics (Table 1). This randomly colored map shows 10 distinct ecoregions identified by the k-means clustering algorithm.

2.3 Ecoregionalization

FLUXNET classifies the sites based on vegetation type they represent using IGBP classification scheme (IGBP, 1990, 1992) among cropland, closed shrubland, deciduous broadleaf forest, evergreen broadleaf forest, evergreen needleleaf forest, grassland, mixed forest, open shrubland, savanna, wetlands and woody savanna. While the IGBP classifications are primarily based on vegetation, a wider range of environmental variables (bioclimatic, edaphic and topographic) were employed in our study. The multivariate clustering technique has been widely used in ecology and Earth sciences for delineation of ecoregions that are relatively homogeneous with respect to selected environmental characteristics (Hargrove and Hoffman, 2005; Bernert et al., 1997; Lark, 1998; Hessburg et al., 2000). A massively parallel k-means clustering analysis tool developed by Kumar et al. (2011) was used to derive custom ecoregions using a selected set of environmental variables (Table 1).

Figure 2 shows the global land area classified into 10 distinct ecoregions at ~ 4 km resolution, when defined quantitatively by the k-means clustering algorithm, using 20 bioclimatic, edaphic and topographic characteristics (Table 1). Quantitatively defined ecoregions correspond well with the major global biomes and serves as a reference landcover classification for this study.

2.4 Representativeness of flux sites

Representativeness of a measurement site is the extent to which the measurements collected at any given location and time represent the conditions at any other location and time. Eddy covariance flux towers measure a range of meteorological and flux variables. While the measurements captured by the sensors are direct measures of a small region in the footprint of the

Commented [R6]: •This appears to be short form of Nappo et al. (1982). Might want to cite.
 •Nappo, C. J., Caneill, J. Y., Furman, R. W., Gifford, F. A., Kaimal, J. C., Kramer, M. L., Lockhart, T. J., Pendergast, M. M., Pielke, R. A., Randerson, D., Shreffler, J. H., and Wyngaard, J. C.: The workshop on the representativeness of meteorological observations, June 1981, Boulder, Colorado, Bull. Am. Meteorol. Soc., 63, 761-764, 1982.



tower, understanding the representativeness of the measurements for a broader landscape is critical for upscaling of point measurements to the larger area. There has been a number of attempts to assess the representativeness of flux tower networks. Hargrove et al. (2003) analyzed the representativeness of the AmeriFlux network using an ecoregion based approach. For each ecoregion in the map, they calculated the Euclidean distance in their data space to the single closest ecoregion that contains a site from the network. Yang et al. (2008) studied the representativeness of the AmeriFlux network using MODIS and GOES data and used Euclidean distance to quantify the similarity between any location and the collection of AmeriFlux sites. Using a combination of footprint modeling and moderate-resolution remote sensing, Chen et al. (2011) calculated the representativeness with a sensor location bias metric. Sulkava et al. (2011) developed a tool to evaluate the representativeness of the European eddy covariance network. They calculated representativeness as the Euclidean distance measured in data space between the cluster centers and the established stations. He et al. (2015) studied the ChinaFlux sites and calculated the representativeness of flux tower sites as the Euclidean distance in a 4-dimensional environmental space. They calculated the Euclidean distance from each pixel to each tower and selected the minimum distance as their network representativeness. Hoffman et al. (2013) utilized climate and permafrost characteristics and derived a Euclidean-based representativeness for a sampling design in the Arctic environment. They developed two approaches for calculating representativeness for sampling network design. First, the ecoregions-based representativeness, where the dissimilarity of any ecoregion containing a sampling site to any other ecoregion, was calculated as the Euclidean distance between the two ecoregion centroids within the standardized n-dimensional data space. The second was a point-based representativeness approach in which the Euclidean distance of each map pixel containing a site from every other pixel on the map was calculated in the standardized n-dimensional data space. When a network of sites is present, the Euclidean distance for the site closest to the pixel in data space was selected as the representativeness of that pixel.

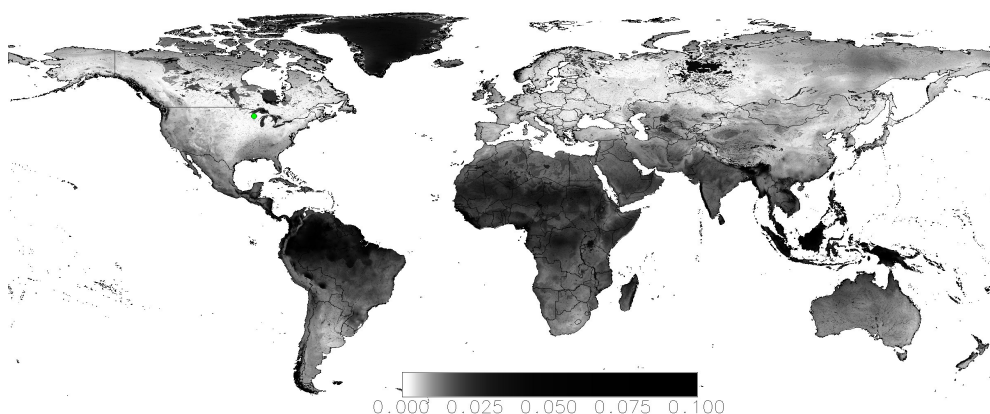
In this study, we used the point-based representativeness approach (Hoffman et al., 2013) to compute the dissimilarity of environmental conditions between any flux site location and every land pixel on the globe at ~4 km resolution in 20-dimensional data space (Table 1). At each flux site the environmental characteristics (Table 1) were extracted from the gridded data sets and the Euclidean distance to every land pixel on the globe was computed as
$$d_{V_{site}, V_{pixel}} = \sqrt{\sum_{n=1}^{20} (V_{site,n} - V_{pixel,n})^2},$$
 resulting in a map showing dissimilarity in environmental conditions sampled by the flux site location to that of other locations on the globe. For example, a representativeness map for the Willow Creek, U.S. site (US-WCr) (Figure 3) located in an upland Deciduous Broadleaf Forest (DBF) shows that the site is representative of temperate environment in central United States and Europe. However, this site does not capture well the dry environment of the southwestern U.S., coastal regions of the southern U.S., evergreen forests in the northwestern U.S. Tropical and high latitude regions of the globe experience significantly different environmental conditions and thus show high dissimilarities when compared to the US-WCr site. Similarly, the Evergreen Broadleaf Forest (EBF) site at Santarum km83, Brazil (BR-Sa3) (Figure 3) represents tropical environments well, but shows high dissimilarities with temperate and high latitude regions.

Environmental conditions at any location on the map may be characteristic of one or many or none of the sites in the sampling network. To understand the best representative from the network as a whole, every pixel was assigned the representativeness corresponding to the site closest to it in the multivariate environmental data space. Figure 4 shows the best representativeness of

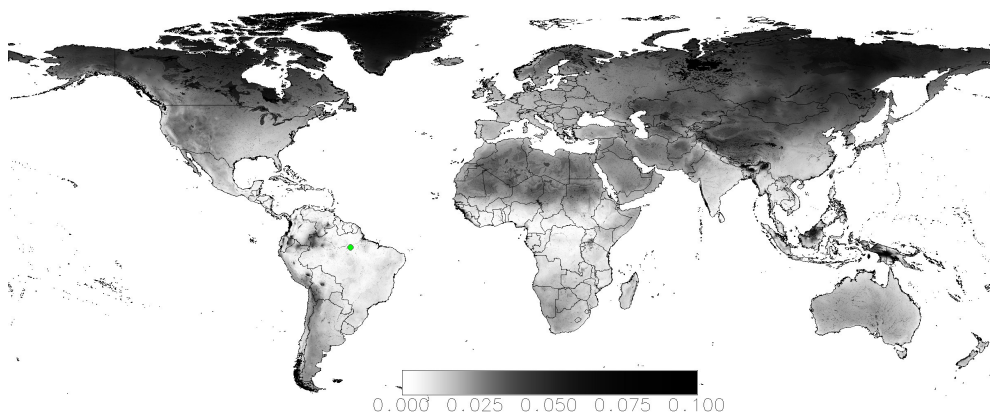
Commented [R7]: •Which varies itself in space over time. Is this transient footprint bias taken into consideration here? If not an estimate of its impact on the upscaling exercise would be very helpful.

Commented [R8]: •It would help to present the approaches in order of the spatial scale they address, and explain their differing objectives on this basis.
 •Please also see comment in introduction section.

Commented [R9]: •Here I get it: A proxy based on multi-dimensional space is used to determine the flux network representativeness for global coverage.
 •I would help the reader to define this more clearly already in the introduction.
 •The question that remains is: how close is close enough, i.e. what uncertainty is incurred for pixels that have a less close site than others?



(a) Willow Creek, U.S. (US-WCr)



(b) Santarum km83, Brazil (BR-Sa3)

Figure 3. Representativeness of individual flux sites.

Commented [R10]: •It would be good to provide sufficient information in the legend so the figure can be interpreted without having to scour the text for relevant information.

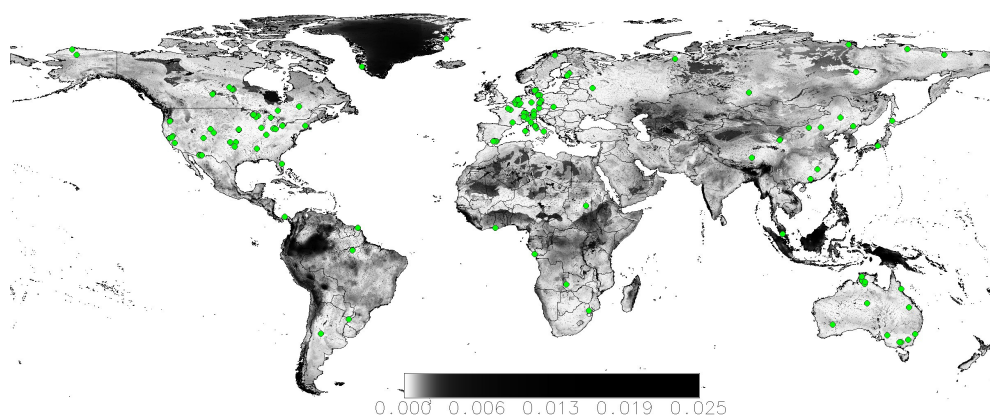


Figure 4. Network representativeness for all of the FLUXNET2015 sites (164 sites).

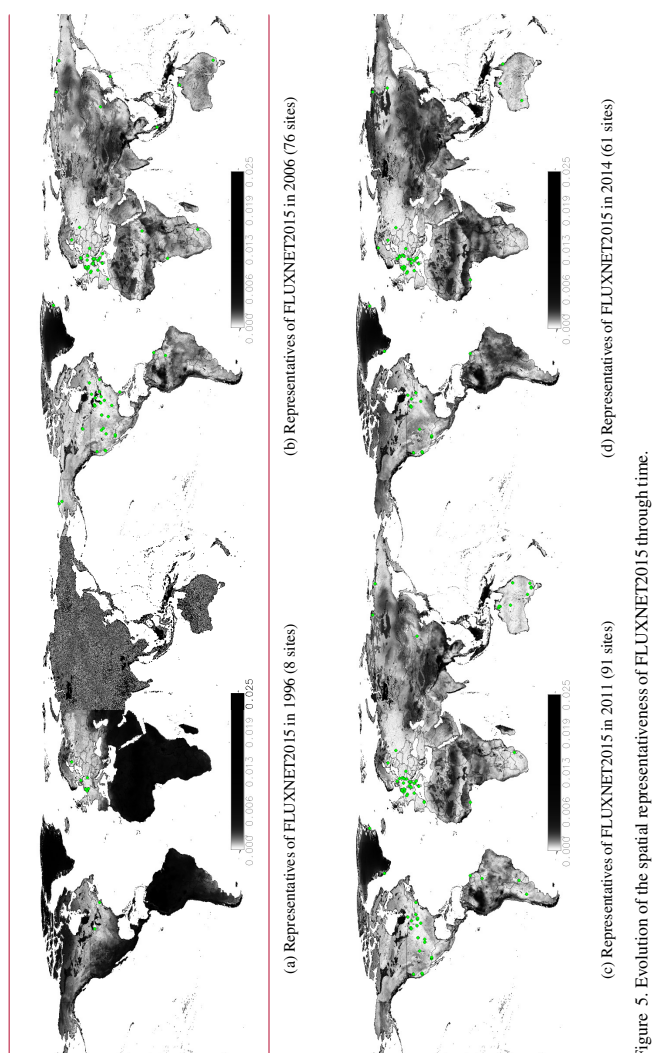
the FLUXNET network considering all 164 sites from FLUXNET2015. FLUXNET2015 shows fairly good representativeness for the large parts of the world except for northern high latitude regions like Iceland, wet tropical forests in South America and southeast Asia.

However, Figure 4 is only the theoretical representativeness of 164 sites in FLUXNET2015, since not all sites are in active operation or have data available each year. As the sites come and goes out of operation, and depending on if data is available or not, our sampling (and thus upscaled view) of terrestrial carbon flux is variable through time. We calculated the representativeness of the network every month considering which sites were active and had data available. Figure 5 shows snapshots of representativeness of the network at different times during the period 1996–2014. Global carbon fluxes were poorly sampled in 1996, when only eight sites were available globally, but has steadily improved since then as sampling sites have been added.

Many eddy covariance measurement sites were added in mid-latitude regions of Europe and North America and Australia, where most of the flux sites are now clustered. Compared to 2011 (Figure 5(c)), 2014 (Figure 5(d)) shows a significant decline in the representativeness of the network globally due to data from a fewer number of sites being available. While many of the sites may be in active operation during recent years, the data from them most probably have not been made available, increasing uncertainties of global, upscaled estimates of carbon flux and highlighting the importance of making data available quickly.

Commented [R11]: •It would be good to provide sufficient information in the legend so the figure can be interpreted without having to scour the text for relevant information.

Deleted: ed



Commented [R12]: •It would be good to provide sufficient information in the legend so the figure can be interpreted without having to scour the text for relevant information.



2.5 Upscaling point measurements

Quantitative representativeness of network sites can be utilized to upscale flux estimates to larger spatial and temporal scales for input to or evaluation of process modeling or for estimating landscape-scale characteristics (Hoffman et al., 2013). Upscaling eddy covariance constrained estimates of gross primary production (GPP) from a global network of flux sites requires 1) identifying all sites that sample environments similar to any location, and 2) developing a statistical model to estimate GPP at all global land regions using estimates from point measurements at flux sites.

2.5.1 Inverse distance weighted interpolation

Interpolation is a process of using measurements about a process at a limited number of point locations to make estimates about the process at other, unmeasured locations. Inverse distance weighting (IDW) is a deterministic, nonlinear interpolation technique that uses a weighted average of the measurements from nearby sample points to estimate the magnitude of variables at non-sampled locations. IDW interpolation is based on Tobler's first law of geography (Tobler, 1970) which states "everything is related to everything else, but near things are more related than distant things". The weight of a measurement at a particular point is assigned in the averaging calculation depending upon the sampled point's distance to the non-sampled location. While traditional spatial interpolation approaches often employ the distance in geographic space between sampled points and the un-sampled location, in this study the Euclidean distance in multivariate environmental data space (representativeness of the sampling location) was used to calculate the weights. Thus, samples from locations closer in terms of environmental conditions were assigned higher weights in interpolation while information about geographical proximity was never used in the process.

An important aspect of the inverse distance weighting interpolation is the neighborhood of influence that determines the sampling points that are used for interpolating the value at any un-sampled location. In this study we systematically defined the neighborhood of influence (described in Section 2.5.2) and thus identified the flux measurements to include in the interpolation at each spatial location and at each monthly time step during the study period.

High temporal resolution measurements at eddy covariance sites capture the temporal dynamics and variability (seasonal and interannual), allowing us to examine the influences of phenology, drought, heat waves, El Niño, length of growing season, presence/absence of snow on canopy-scale fluxes, etc. (Baldocchi et al., 2001). Statistical models built using all the data, while representing the long term trends and mean climatology well, also tend to lose the fine temporal dynamics captured by measurements. To preserve the rich temporal dynamics in the flux measurements, we built the interpolation at monthly time steps using the data only from that month. This approach, distinct from many past studies, helps capture the temporal variability in the upscaled data set; however, it is also prone to exhibit a bias if a localized phenomena is experienced by a flux site.

2.5.2 Identifying flux sites in similar environments

While global terrestrial ecosystems are highly diverse and heterogeneous, point measurements at flux sites sample the environment in which they are located. These measurements are representative of similar conditions at geographic locations elsewhere. The network of global flux sites which are managed and operated by independent regional institutions provide a sparse and

Commented [R13]: •It is my understanding that such approach does not take into account inter-annual differences in environmental conditions (the proxy used to determine Euclidean distance) across space.
 •If this is the case, how can a defensible time-resolved surface flux product be created if the mapping/interpolation operator is constant in time?
 •This poses the question, why does the study not also employ time-resolved satellite data products to fill this gap?
 •Also, for pixels with state-space combinations that are observed by many flux sites, the resulting uncertainty should be smaller compared to those observed by few flux sites.
 •Hence, I think an uncertainty budget is indispensable for determining the accuracy and usefulness of the presented approach.

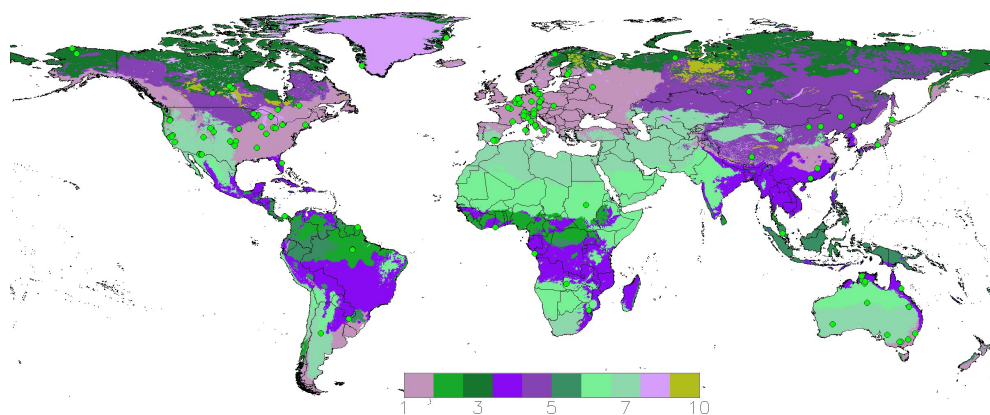


Figure 6. Spatial distribution of FLUXNET sites across global ecoregions.

non-uniform sampling of the environment. While some regions and environments may be well sampled, others may be undersampled. Thus it is important to identify the flux sites in similar environments where GPP estimates were available for upscaling. We overlaid the locations of sites in FLUXNET over the quantitatively defined ecoregions (Figure 2) and identified the sites in each ecoregion. Figure 6 shows the distribution of 164 FLUXNET sites in the FLUXNET2015 collection globally across various ecoregions. To estimate the GPP at any location, the sites located within the same ecoregion would provide the most relevant and representative measurements. Ecoregions in the mid-latitudes of North America and Europe are well sampled by the global network of flux sites, while sites in many of the other ecoregions are few to none. Figure 4 shows the best possible representativeness while using all 164 sites in the FLUXNET2015 collection. The actual availability of flux site measurements is spatio-temporally variable.

Table 2 shows the number of sites within each ecoregion each year during the period 1991–2014 for which data were available in the FLUXNET2015 collection. We carefully processed and identified the sites at which GPP estimates were available within each ecoregion monthly for use in the upscaled product. However, there were some ecoregions and time periods that were completely unsampled by the available sites. In such cases we identified the ecoregion most similar in multivariate data space (Table 1) that was sampled by at least one or more flux sites and used those data for extrapolation. The accuracy of the GPP estimates in such cases is expected to suffer due to data limitation.

3 Results and discussions

We used monthly time series of GPP estimates from eddy covariance observations taken at flux towers from FLUXNET2015 to develop upscaled gridded GPP data sets at ~4 km resolution. We compared the upscaled GPP time series data set to

Commented [R14]: •This figure is redundant and can be combined with Figure 1.

Commented [R15]: •...and requires quantification.



Table 2. Active flux sites present in each ecoregion during 1991–2014. Data from the period 1991–1999 was not used in the study. Some ecoregions had no flux sites available during the period 2000–2014.

Year	Total	Ecoregions									
		1	2	3	4	5	6	7	8	9	10
		14.27%	4.19%	14.65%	12.19%	17.84%	3.12%	12.21%	14.82%	5.45%	1.25%
1991	1	1	0	0	0	0	0	0	0	0	0
1992	1	1	0	0	0	0	0	0	0	0	0
1993	1	1	0	0	0	0	0	0	0	0	0
1994	1	1	0	0	0	0	0	0	0	0	0
1995	2	2	0	0	0	0	0	0	0	0	0
1996	8	8	0	0	0	0	0	0	0	0	0
1997	10	9	1	0	0	0	0	0	0	0	0
1998	13	12	1	0	0	0	0	0	0	0	0
1999	15	14	1	0	0	0	0	0	0	0	0
2000	26	21	2	1	1	0	0	0	1	0	0
2001	40	26	2	3	1	0	0	0	3	0	5
2002	61	35	3	4	2	6	0	0	3	1	7
2003	73	37	3	7	5	8	1	0	3	1	8
2004	90	48	4	8	5	8	1	0	6	2	8
2005	89	49	3	8	5	5	1	1	7	2	8
2006	76	50	3	6	5	2	1	1	6	1	1
2007	80	51	5	5	5	3	1	1	8	1	0
2008	86	53	4	6	7	3	1	2	9	1	0
2009	89	54	4	4	6	3	1	2	14	1	0
2010	84	53	2	4	3	4	0	1	16	1	0
2011	91	57	3	3	3	2	0	3	18	2	0
2012	93	57	2	3	3	3	0	3	19	3	0
2013	85	51	2	3	3	2	0	3	18	3	0
2014	61	38	2	2	1	2	0	1	13	2	0

FLUXNET-MTE (Jung et al., 2009). While FLUXNET-MTE GPP was also developed using observations from FLUXNET sites, a different set of sites and data products were used in the current study. The overlapping period of 2000–2008 was used for comparison with FLUXNET-MTE.

3.1 Spatial distribution of GPP

- 5 Figure 7 shows the global spatial distribution of time integrated mean annual GPP for four years in the comparison period. The quality and accuracy of the upscaled GPP time series data is strongly affected by the availability and spatial distribution of the flux sites at any given time. Figure 8 shows the comparison of our upscaled GPP estimates with FLUXNET-MTE GPP estimates (Jung et al., 2009). Agreement was fairly good for Northern Hemisphere mid-latitudes regions that were well sampled

Commented [R16]: •Where is accuracy being quantified? It would be very helpful to show gridded uncertainty estimates.



by FLUXNET2015 sites. However, the data show significant differences in Southern Hemisphere tropical regions, which lack sufficient flux measurements. For example, FLUXNET-MTE employs the data for the flux sites in Matto Grosso (BR-Mtg) in South America, which was not available as part of FLUXNET2015 data set, leading to an underestimate of GPP in that region compared to FLUXNET-MTE. FLUXNET-MTE also suffers from limitations of data in this region, indicated by their high index of extrapolation in tropical areas (Jung et al., 2009).

Extra-tropical regions in the Southern Hemisphere were poorly sampled by the FLUXNET2015 data with only a few sites making data available in that region. In addition, based on the environmental similarity sites in the Northern Hemisphere were used to estimate the fluxes in the region, which have an out-of-phase phenology, leading to an overestimation of GPP magnitude in Northern Hemisphere winter.

Higher estimates of GPP compared to FLUXNET-MTE in our upscaled data set in coastal regions of the Pacific Northwest of North America provide an improved representation of the wet evergreen forests. Coastal regions of the Pacific Northwest possess of high orographic relief and strong rain shadows. FLUXNET-MTE does not capture these highly productive forests, probably due to lower resolution averaging over wet and dry forests, but our approach at a high resolution provides a better estimate of GPP in this orographically complex terrain.

Commented [R17]: •Nice showcasing of how higher resolution can remedy non-linear aggregation effects.

3.2 Seasonal patterns of GPP

Figure 9 shows the zonal mean seasonal patterns of monthly GPP estimated from FLUXNET2015 for years 2000, 2006, 2011, and 2014. In all cases, June and July exhibit the large GPP in the Northern Hemisphere extra-tropics. Weak seasonality, inter-annual variability, and regional under-sampling results in non-persistent ordering of monthly GPP in the tropics. As mentioned above, Southern Hemisphere high latitude regional estimates are strongly influenced by measurements in Northern Hemisphere mid-latitudes because of the similarity in their environmental conditions and the lack of FLUXNET2015 measurement sites in the Southern Hemisphere. Consequently, anomalously high GPP is estimated during June and July in -30° to -50° latitude. The varying list of measurement sites each year also affects upscaled GPP estimates. For example, many fewer sites have made data from year 2014 available in FLUXNET2015, so Figure 9(c) exhibits unusually low tropical GPP. Figure 10(a) shows significant spread in our annual upscaled GPP because of both interannual variability and as a consequence of the continuously evolving list of contributing measurement sites. However, it is likely that interannual variability is strong enough to result in great variability in GPP than is exhibited by the FLUXNET-MTE estimates in Figure 10(b).

3.3 Temporal variability of GPP

Eddy covariance observations at flux sites capture changes in biosphere–atmosphere exchange at a high temporal resolution. Temporal variability is due to long term changes in environmental conditions, vegetation dynamic, disturbance events, etc. To preserve the temporal variability in GPP captured by the instruments in the upscaled product, only the observations from the same month were used in our IDW-based interpolation method.

Thus the pattern of time integrated annual GPP over the years show significant interannual variability for all latitudinal zones (Figure 10(a)). Variability was especially high in tropical region where only a few flux sites were present in FLUXNET2015

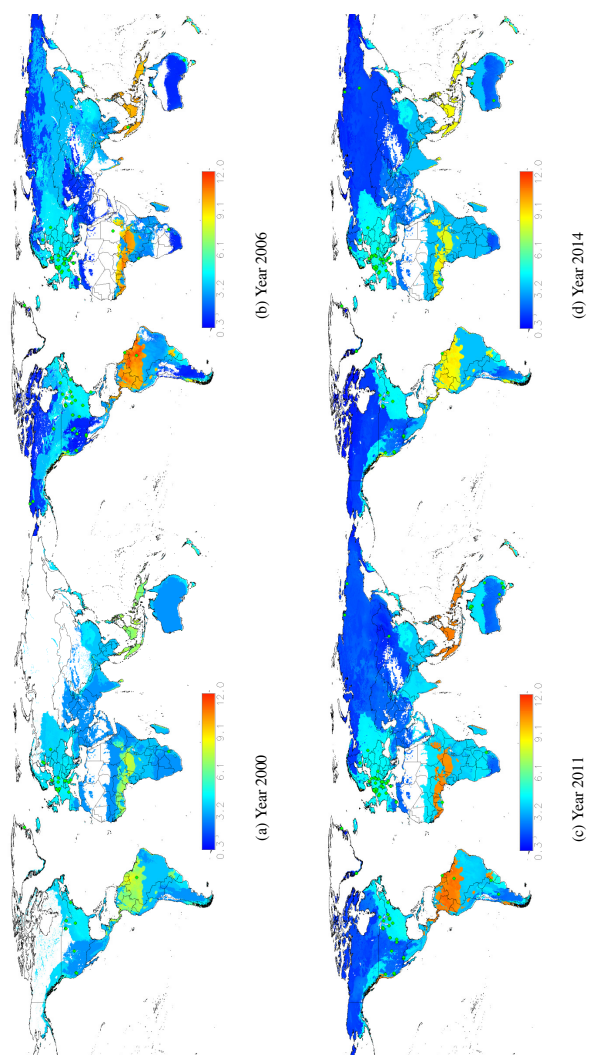


Figure 7. Spatial distribution of time integrated annual mean GPP fluxes [Units: $\text{g Cm}^{-2} \text{d}^{-1}$].

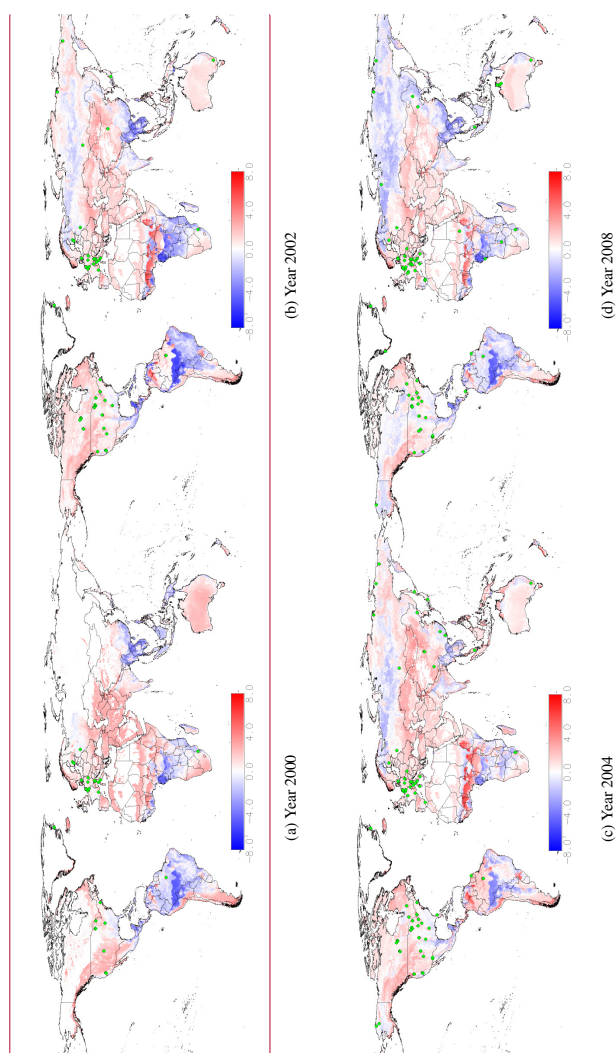


Figure 8. Comparison of upscaled annual mean GPP fluxes to FLUXNET-MTE GPP. Positive values indicate higher estimates, and negative indicate lower estimated GPP compared to FLUXNET-MTE. [Units: $\text{g C m}^{-2} \text{ d}^{-1}$]

Commented [R18]: •Adding a relative scale would be helpful for straightforward interpretability.

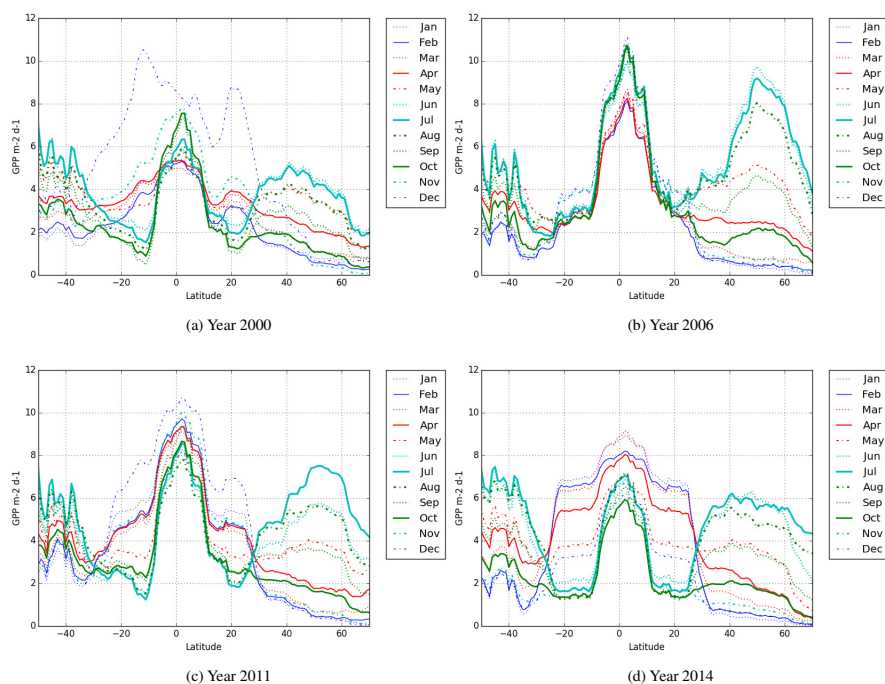


Figure 9. Seasonal patterns of mean monthly GPP across latitudinal bands. [Units: $\text{g C m}^{-2} \text{d}^{-1}$]

data set. The magnitude and temporal variability of the upscaled GPP estimates were thus significantly influenced by regions where flux sites were present and may not be reflective of reality. Vegetation responses to the severe 2005 drought in the Amazon has been a topic of controversy with studies reporting patterns of greening (Saleska et al., 2007) and others reporting decline (Samanta et al., 2010). Our GPP product shows a green up during the 2005–2006 period in tropical zones; however, this greening cannot be fully confirmed because of the lack of sufficient flux sites to adequately sample tropical regions for the spatio-temporal pattern of vegetation response to drought. For example, Santarem Km67 site (BR-Sa1) in Tapajos National Forest in Amazonia Brazil was one of the few sites in the tropics that was available in FLUXNET2015. Patterns of GPP observations at BR-Sa1 (Figure 11), including a greening trend during 2005–2006 drought, strongly influenced our upscaled data set. Flux measurements from more sites, including existing sites which are not available within FLUXNET2015, would help improve the upscaled estimates of GPP.

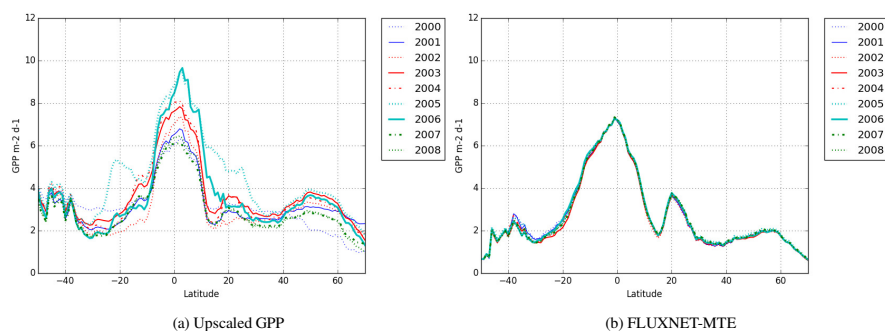


Figure 10. Interannual variability in GPP fluxes for the period 2000–2008. [Units: $\text{g C m}^{-2} \text{d}^{-1}$]

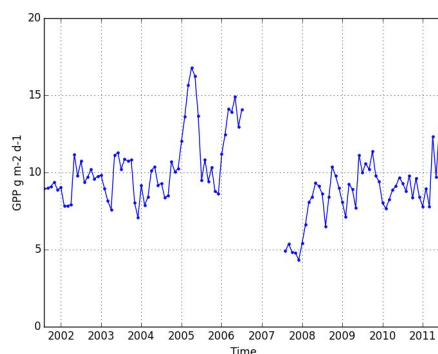


Figure 11. Observed GPP timeseries at Santarem Km67 site (BR-Sa1), Brazil [Units: $\text{g C m}^{-2} \text{d}^{-1}$]

While FLUXNET-MTE shows a similar latitudinal pattern, interannual variability was significantly lower (Figure 10(b)). The FLUXNET-MTE model was trained using data across all years, and thus possibly was less susceptible to the strong interannual variability observed at the flux sites and rather is closer to a climatological mean pattern.

Due to variability in the number and spatial distribution of flux sites for which data were available in FLUXNET2015, the upscaled GPP product also exhibits year to year variability in accuracy. Figure 12 and Table 3 summarize the statistical comparison of the upscaled GPP with FLUXNET-MTE GPP as reference benchmark.

Commented [R19]: •I am not sure, because the mapping operator in the present manuscript appears to be driven by proxies on a climatic timescale, i.e. without resolved inter-annual variability.

•What however does introduce inter-annual variability in the current study is that sites informing the mapping operation are being switched on/off depending on data availability in the given month.

•So it remains unclear whether the larger variability is a natural phenomenon, or an artifact of the mapping operation.

•I suggest that the susceptibility of the proposed mapping operation be studied and quantified in an uncertainty budget which will allow to substantiate the provided interpretation.

Commented [R20]: •Requires further explanation, cannot be interpreted with the little text provided here and in the legends.

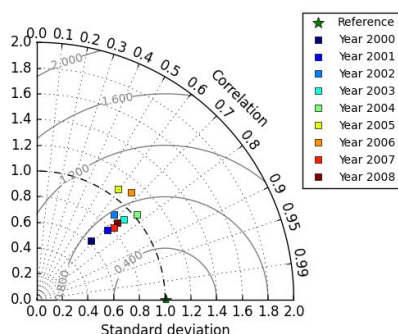


Figure 12. Statistical comparison of upscaled time integrated annual mean GPP with FLUXNET-MTE GPP.

Table 3. Comparison statistics for upscaled GPP data set compared to FLUXNET-MTE

Year	RMSE	Correlations (R)
2000	1.959	0.675
2001	1.557	0.717
2002	1.717	0.669
2003	1.710	0.739
2004	1.652	0.763
2005	2.425	0.594
2006	1.974	0.661
2007	1.484	0.730
2008	1.512	0.724

4 Conclusions

Networks of global flux sites capture the dynamic terrestrial biosphere–atmosphere gas exchange and provides us insights into terrestrial ecosystem processes. In this study we have presented a method to assess the representativeness of the global network of flux sites, specifically those included in the FLUXNET2015 data set. Due to independent operation of member flux sites in the FLUXNET network, a wider range of variation exists in duration of operation, measurements being conducted and availability of data sets for open research. Thus unlike previous studies, we conducted our assessment to quantify the representativeness of the network through time. The analysis provides a window into the evolution of the global eddy covariance measurements over time, quantify the contribution of any individual sites, identify the regions that are better represented by addition of sites over time, and ones that continue to be under-represented.

Upscaling of measured carbon fluxes is key for exploiting the full potential of rich FLUXNET2015 observations beyond the tower footprints to landscape scale understanding of ecosystem processes. We also developed a representativeness-based

Commented [R21]: •It would be good to provide sufficient information in the figure axes and legend so the figure can be interpreted without having to scour the text for relevant information.

Commented [R22]: •I agree, this is the strong suit of this study.



upscaling approach to develop monthly time series high resolution global gridded estimates of GPP. Our upscaling method utilized the observations only from flux sites that sampled similar environmental conditions. Spatio-temporal variability and accuracy of the upscaled GPP data set strongly reflects the spatio-temporal variability in distribution of flux observations. Thus, the data set has higher accuracy and low uncertainty in the ecoregions that are well sampled by the available flux sites, and lower accuracy and higher uncertainty in the ecoregions where available observations are sparse to none. This study was an attempt to understand the global carbon cycle (specifically GPP) through the lens of eddy covariance measurements across the globe. This work highlights the regions that are poorly understood by the existing (and available) networks, like Southern Hemisphere in general and carbon rich tropical forest ecosystems in particular.

This approach can easily be applied to flux data as they become available in order to routinely derive upscaled carbon fluxes.

While the current study was focused on GPP fluxes, the same method can be applied to other carbon, water and energy fluxes.

While a larger number of sites are part of the full global FLUXNET network, full potential of the network to understand the global carbon cycle can be realized only when data from all the sites are made available for research and analysis. Our future efforts will focus on integration of data sets from other flux and forest inventory networks, like the Global Ecosystem Monitoring (GEM) and RAINFOR networks, especially to improve our understanding of biodiversity and ecosystem function in carbon rich tropical forests.

Data availability

All data sets from this study have been archived (Kumar et al., 2016) as a part of Next Generation Ecosystem Experiments (NGEE) Tropics Collection of Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory and are publicly available at <http://dx.doi.org/10.15486/NGT/1279968>. All gridded data are distributed in a CF Compliant NetCDF format.

Author contributions. JK, FMH, WWH designed the study and developed the method to quantify network representativeness. JK processed the data, developed analysis method/tools and upscaled data sets. JK, FMH, WWH and NC analyzed the data. JK led the writing of manuscript with contributions from all the authors.

Acknowledgements. This work was supported by the Next Generation Ecosystem Experiments Tropics (NGEE-Tropics) and ORNL Terrestrial Ecosystem Science Scientific Focus Area projects, sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe>).

Commented [R23]: •There is not a single quantitative accuracy metric to be found in the manuscript, so I don't think this is a valid conclusion.

Commented [R24]: •Please quantify, are we talking 1%, 10%, 100%?

Deleted: S

Commented [R25]: •This would make for a great sentence in the introduction. There it could be combined with the explicit notion that flux measurements are only combined with a mapping operator based on climatological state-space representations to arrive at gridded, global data product. And that one main difference to other studies is that no additional time-resolved proxies such as satellite data are being used.



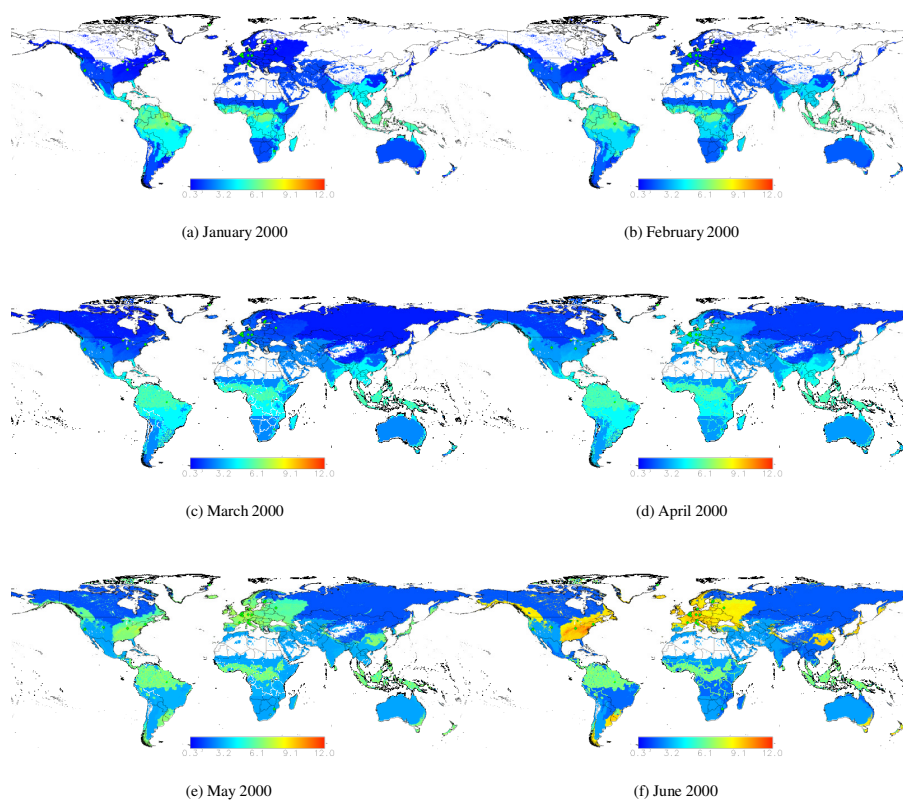
public-access-plan). This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

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 FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices.



Appendix A: Upscaled monthly GPP time series

Figure 13 shows the upscaled monthly time series of spatial distribution of global GPP fluxes for the year 2000.



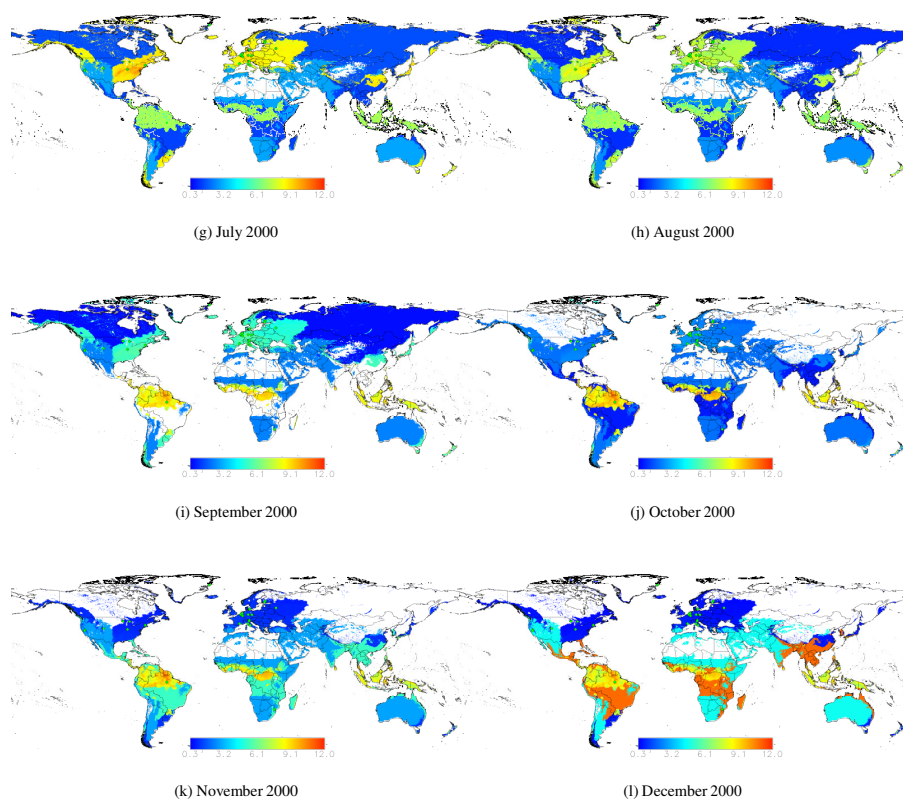


Figure 13. Spatial distribution of upscaled global monthly GPP for year 2000.



References

- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, *Bulletin of the American Meteorological Society*, 82, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, [http://dx.doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](http://dx.doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2), 2001.
- Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, *Global Change Biology*, 9, 479–492, doi:10.1046/j.1365-2486.2003.00629.x, <http://dx.doi.org/10.1046/j.1365-2486.2003.00629.x>, 2003.
- Bernert, A. J., Eilers, M. J., Freemark, E. K., and Ribic, C.: A Quantitative Method for Delineating Regions: An Example for the Western Corn Belt Plains Ecoregion of the USA, *Environmental Management*, 21, 405–420, <http://dx.doi.org/10.1007/s002679900038>, 1997.
- Chapin, F. Stuart, I., Matson, P. A., and Vitousek, P. M.: *Principles of Terrestrial Ecosystem Ecology*, Springer New York, doi:10.1007/978-1-4419-9504-9, 2012.
- Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., Black, T. A., Arain, M. A., Bourque, C. P.-A., Flanagan, L. B., Lafleur, P. M., McCaughey, J. H., and Wofsy, S. C.: Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling, *Agricultural and Forest Meteorology*, 151, 87 – 100, doi:<http://dx.doi.org/10.1016/j.agrformet.2010.09.005>, <http://www.sciencedirect.com/science/article/pii/S0168192310002510>, 2011.
- Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Black, T. A., Arain, M. A., Barr, A. G., Bourque, C. P.-A., Flanagan, L. B., Lafleur, P. M., McCaughey, J. H., and Wofsy, S. C.: 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, doi:<http://dx.doi.org/10.1016/j.rse.2012.06.007>, <http://www.sciencedirect.com/science/article/pii/S003442571200243X>, 2012.
- Global Soil Data Task Group: Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System)]., doi:10.3334/ORNLDAC/569, <http://dx.doi.org/10.3334/ORNLDAC/569>, 2000.
- Hargrove, W. W. and Hoffman, F. M.: Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions, *Environ. Manage.*, 34, S39–S60, doi:10.1007/s00267-003-1084-0, 2005.
- Hargrove, W. W., Hoffman, F. M., and Law, B. E.: New analysis reveals representativeness of the AmeriFlux network, *Eos, Transactions American Geophysical Union*, 84, 529–535, doi:10.1029/2003EO480001, <http://dx.doi.org/10.1029/2003EO480001>, 2003.
- He, H., Zhang, L., Gao, Y., Ren, X., Zhang, L., Yu, G., and Wang, S.: Regional representativeness assessment and improvement of eddy flux observations in China, *Science of The Total Environment*, 502, 688–698, <http://www.sciencedirect.com/science/article/pii/S0048969714013953>, 2015.
- Hessburg, P., Salter, R., Richmond, M., and Smith, B.: Ecological subregions of the Interior Columbia Basin, USA, *Applied Vegetation Science*, 3, 163–180, doi:10.2307/1478995, <http://dx.doi.org/10.2307/1478995>, 2000.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology*, 25, 1965–1978, doi:10.1002/joc.1276, <http://dx.doi.org/10.1002/joc.1276>, 2005.



Hoffman, F. M., Kumar, J., Mills, R. T., and Hargrove, W. W.: Representativeness-based Sampling Network Design for the State of Alaska, *Landscape Ecol.*, 28, 1567–1586, doi:10.1007/s10980-013-9902-0, 2013.

IGBP: The International Geosphere–Biosphere Programme: a study of global change—the initial core project, IGBP Global Change Report no. 12, International Geosphere–Biosphere Programme, Stockholm, Sweden, 1990.

5 IGBP: Improved Global Data for Land Applications, edited by J. R. G. Townshend. IGBP Global Change Report no. 20, International Geosphere–Biosphere Programme, Stockholm, Sweden, 1992.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, *Biogeosciences*, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, <http://www.biogeosciences.net/6/2001/2009/>, 2009.

10 Kumar, J., Mills, R. T., Hoffman, F. M., and Hargrove, W. W.: Parallel k-Means Clustering for Quantitative Ecoregion Delineation Using Large Data Sets, in: *Proceedings of the International Conference on Computational Science (ICCS 2011)*, edited by Sato, M., Mat-suoka, S., Sloot, P. M., van Albada, G. D., and Dongarra, J., vol. 4 of *Procedia Comput. Sci.*, pp. 1602–1611, Elsevier, Amsterdam, doi:10.1016/j.procs.2011.04.173, 2011.

Kumar, J., Hoffman, F. M., Hargrove, W. W., and Collier, N.: Global 4 km resolution monthly gridded Gross Primary Productivity (GPP) data set derived from FLUXNET2015 [Data], Next Generation Ecosystem Experiments Tropics Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.15486/NGT/1279968, 2016.

Lark, R. M.: Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield, *International Journal of Geographical Information Science*, 12, 83–98, doi:10.1080/136588198242021, <http://dx.doi.org/10.1080/136588198242021>, 1998.

20 Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneeth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global Change Biology*, 16, 187–208, doi:10.1111/j.1365-2486.2009.02041.x, <http://dx.doi.org/10.1111/j.1365-2486.2009.02041.x>, 2010.

Pastorello, G., Agarwal, D., Papale, D., Samak, T., Trotta, C., Ribeca, A., Poindexter, C., Faybishenko, B., Gunter, D., Hollowgrass, R., and Canfora, E.: Observational Data Patterns for Time Series Data Quality Assessment, in: *e-Science (e-Science)*, 2014 IEEE 10th International Conference on, vol. 1, pp. 271–278, doi:10.1109/eScience.2014.45, 2014.

25 Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, <http://dx.doi.org/10.1111/j.1365-2486.2005.001002.x>, 2005.

Saleska, S. R., Didan, K., Huete, A. R., and da Rocha, H. R.: Amazon Forests Green-Up During 2005 Drought, *Science*, 318, 612–612, doi:10.1126/science.1146663, <http://science.sciencemag.org/content/318/5850/612>, 2007.

35 Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., Nemani, R. R., and Myneni, R. B.: Amazon forests did not green-up during the 2005 drought, *Geophysical Research Letters*, 37, n/a–n/a, doi:10.1029/2009GL042154, <http://dx.doi.org/10.1029/2009GL042154>, 105401, 2010.

Saxon, E., Baker, B., Hargrove, W., Hoffman, F., and Zganjar, C.: Mapping Environments at Risk Under Different Global Climate Change Scenarios, *Ecol. Lett.*, 8, 53–60, doi:10.1111/j.1461-0248.2004.00694.x, 2005.



- Schlesinger, W. H. and Bernhardt, E. S.: Biogeochemistry (Third Edition), pp. 275–340, Academic Press, Boston, doi:10.1016/B978-0-12-385874-0.00008-X, <http://www.sciencedirect.com/science/article/pii/B978012385874000008X>, 2013.
- Sulkava, M., Luyssaert, S., Zachle, S., and Papale, D.: Assessing and improving the representativeness of monitoring networks: The European flux tower network example, *Journal of Geophysical Research: Biogeosciences*, 116, n/a–n/a, doi:10.1029/2010JG001562, <http://dx.doi.org/10.1029/2010JG001562>, g00J04, 2011.
- 5 Tobler, W. R.: A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46, 234–240, 1970.
- Yang, F., Zhu, A.-X., Ichii, K., White, M. A., Hashimoto, H., and Nemani, R. R.: Assessing the representativeness of the AmeriFlux network using MODIS and GOES data, *Journal of Geophysical Research: Biogeosciences*, 113, n/a–n/a, doi:10.1029/2007JG000627, <http://dx.doi.org/10.1029/2007JG000627>, g04036, 2008.