

## Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites

Housen Chu<sup>a,\*</sup>, Xiangzhong Luo<sup>a,bf</sup>, Zutao Ouyang<sup>b</sup>, W. Stephen Chan<sup>a</sup>, Sigrid Dengel<sup>a</sup>, Sébastien C. Biraud<sup>a</sup>, Margaret S. Torn<sup>a</sup>, Stefan Metzger<sup>c,d</sup>, Jitendra Kumar<sup>e</sup>, M. Altaf Arain<sup>f</sup>, Tim J. Arkebauer<sup>g</sup>, Dennis Baldocchi<sup>h</sup>, Carl Bernacchi<sup>i</sup>, Dave Billesbach<sup>g</sup>, T. Andrew Black<sup>j</sup>, Peter D. Blanken<sup>k</sup>, Gil Bohrer<sup>l</sup>, Rosvel Bracho<sup>m</sup>, Shannon Brown<sup>n</sup>, Nathaniel A. Brunsell<sup>o</sup>, Jiquan Chen<sup>p</sup>, Xingyuan Chen<sup>q</sup>, Kenneth Clark<sup>r</sup>, Ankur R. Desai<sup>d</sup>, Tomer Duman<sup>s</sup>, David Durden<sup>c</sup>, Silvano Fares<sup>t</sup>, Inke Forbrich<sup>u</sup>, John A. Gamon<sup>v,w</sup>, Christopher M. Gough<sup>x</sup>, Timothy Griffis<sup>y</sup>, Manuel Helbig<sup>z,az</sup>, David Hollinger<sup>aa</sup>, Elyn Humphreys<sup>ab</sup>, Hiroki Ikawa<sup>ac</sup>, Hiroki Iwata<sup>ad</sup>, Yang Ju<sup>l</sup>, John F. Knowles<sup>ae</sup>, Sara H. Knox<sup>af</sup>, Hideki Kobayashi<sup>ag</sup>, Thomas Kolb<sup>bg</sup>, Beverly Law<sup>aj</sup>, Xuhui Lee<sup>ak</sup>, Marcy Litvak<sup>s</sup>, Heping Liu<sup>al</sup>, J. William Munger<sup>am</sup>, Asko Noormets<sup>an</sup>, Kim Novick<sup>ao</sup>, Steven F. Oberbauer<sup>ap</sup>, Walter Oechel<sup>aq</sup>, Patty Oikawa<sup>ar</sup>, Shirley A. Papuga<sup>as</sup>, Elise Pendall<sup>at</sup>, Prajaya Prajapati<sup>an</sup>, John Prueger<sup>au</sup>, William L Quinton<sup>av</sup>, Andrew D. Richardson<sup>ah,ai</sup>, Eric S. Russell<sup>al</sup>, Russell L. Scott<sup>ae</sup>, Gregory Starr<sup>aw</sup>, Ralf Staebler<sup>ax</sup>, Paul C. Stoy<sup>d</sup>, Ellen Stuart-Haentjens<sup>ay</sup>, Oliver Sonnentag<sup>az</sup>, Ryan C. Sullivan<sup>ba</sup>, Andy Suyker<sup>g</sup>, Masahito Ueyama<sup>bb</sup>, Rodrigo Vargas<sup>bc</sup>, Jeffrey D. Wood<sup>bd</sup>, Donatella Zona<sup>aq,be</sup>

<sup>a</sup> Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>b</sup> Department of Earth System Science, Stanford University, Stanford, CA, USA

<sup>c</sup> National Ecological Observatory Network Program, Battelle, Boulder, CO, USA

<sup>d</sup> Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, USA

<sup>e</sup> Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

<sup>f</sup> School of Geography and Earth Sciences, McMaster University, Ontario, Canada

<sup>g</sup> Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, USA

<sup>h</sup> Department of Environmental Sciences, Policy, and Management, University of California, Berkeley, CA, USA

<sup>i</sup> USDA-ARS Global Change and Photosynthesis Research Unit

<sup>j</sup> Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, Canada

<sup>k</sup> Department of Geography, University of Colorado, Boulder, CO, USA

<sup>l</sup> Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, OH, USA

<sup>m</sup> School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA

<sup>n</sup> School of Environmental Sciences, University of Guelph, Guelph, Ontario, Canada

<sup>o</sup> Department of Geography & Atmospheric Science, University of Kansas, Lawrence, KS, USA

<sup>p</sup> Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI, USA

<sup>q</sup> Earth System Measurements & Data, Pacific Northwest National Laboratory, Richland, WA, USA

<sup>r</sup> USDA Forest Service, Northern Research Station, New Lisbon, NJ, USA

<sup>s</sup> Department of Biology, University of New Mexico, Albuquerque, NM, USA

<sup>t</sup> National Research Council, Institute of BioEconomy, Rome, Italy

<sup>u</sup> Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA, USA

<sup>v</sup> Department of Earth and Atmospheric Sciences and Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

<sup>w</sup> School of Natural Resources, University of Nebraska - Lincoln, NE, USA

<sup>x</sup> Department of Biology, Virginia Commonwealth University, Richmond, VA, USA

<sup>y</sup> Department of Soil, Water, and Climate, University of Minnesota, Saint Paul, MN, USA

<sup>z</sup> Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>aa</sup> Northern Research Station, USDA Forest Service, Durham, NH, USA

<sup>ab</sup> Department of Geography and Environmental Studies, Carleton University, Ottawa, Ontario, Canada

<sup>ac</sup> Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization, Tsukuba, Japan

<sup>ad</sup> Department of Environmental Science, Shinshu University, Matsumoto, Japan

\* Corresponding author.

<sup>ae</sup> Southwest Watershed Research Center, USDA-Agricultural Research Service, Tucson, AZ, USA  
<sup>af</sup> Department of Geography, The University of British Columbia, Vancouver, British Columbia, Canada  
<sup>ag</sup> Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan  
<sup>ah</sup> Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ 86011, USA  
<sup>ai</sup> School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA  
<sup>aj</sup> Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR, USA  
<sup>ak</sup> School of the Environment, Yale University, New Haven, CT, USA  
<sup>al</sup> Department of Civil and Environmental Engineering, Washington State University, Pullman, WA, USA  
<sup>am</sup> School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA  
<sup>an</sup> Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA  
<sup>ao</sup> O’Neill School of Public and Environmental Affairs, Indiana University – Bloomington, Bloomington, IN, USA  
<sup>ap</sup> Department of Biological Sciences, Florida International University, Miami, FL, USA  
<sup>aq</sup> Department of Biology, San Diego State University, San Diego, CA, USA  
<sup>ar</sup> Department of Earth & Environmental Sciences, California State University, East Bay, Hayward, CA, USA  
<sup>as</sup> Department of Environmental Science and Geology, Wayne State University, Detroit, MI, USA  
<sup>at</sup> Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, Australia  
<sup>au</sup> USDA-Agricultural Research Service, National Laboratory for Agriculture and the Environment, Ames, IA, USA  
<sup>av</sup> Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, Ontario, Canada  
<sup>aw</sup> Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, USA  
<sup>ax</sup> Environment and Climate Change Canada, Toronto, Ontario, Canada  
<sup>ay</sup> California Water Science Center, United States Geological Survey, Sacramento, CA, USA  
<sup>az</sup> Département de géographie, Université de Montréal, Montréal, Québec, Canada  
<sup>ba</sup> Environmental Science Division, Argonne National Laboratory, Lemont, IL, USA  
<sup>bb</sup> Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Japan  
<sup>bc</sup> Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware, USA  
<sup>bd</sup> School of Natural Resources, University of Missouri, Columbia, MO, USA  
<sup>be</sup> Department of Animal and Plant Science, University of Sheffield, UK  
<sup>bf</sup> Department of Geography, National University of Singapore, Singapore  
<sup>bg</sup> School of Forestry, Northern Arizona University, Flagstaff, AZ, USA

## ARTICLE INFO

**Keywords:**  
 Flux footprint  
 Spatial representativeness  
 Landsat EVI  
 Land cover  
 Sensor location bias  
 Model-data benchmarking

## ABSTRACT

Large datasets of greenhouse gas and energy surface-atmosphere fluxes measured with the eddy-covariance technique (e.g., FLUXNET2015, AmeriFlux BASE) are widely used to benchmark models and remote-sensing products. This study addresses one of the major challenges facing model-data integration: To what spatial extent do flux measurements taken at individual eddy-covariance sites reflect model- or satellite-based grid cells? We evaluate flux footprints—the temporally dynamic source areas that contribute to measured fluxes—and the representativeness of these footprints for target areas (e.g., within 250–3000 m radii around flux towers) that are often used in flux-data synthesis and modeling studies. We examine the land-cover composition and vegetation characteristics, represented here by the Enhanced Vegetation Index (EVI), in the flux footprints and target areas across 214 AmeriFlux sites, and evaluate potential biases as a consequence of the footprint-to-target-area mismatch. Monthly 80% footprint climatologies vary across sites and through time ranging four orders of magnitude from  $10^3$  to  $10^7$  m<sup>2</sup> due to the measurement heights, underlying vegetation- and ground-surface characteristics, wind directions, and turbulent state of the atmosphere. Few eddy-covariance sites are located in a truly homogeneous landscape. Thus, the common model-data integration approaches that use a fixed-extent target area across sites introduce biases on the order of 4%–20% for EVI and 6%–20% for the dominant land cover percentage. These biases are site-specific functions of measurement heights, target area extents, and land-surface characteristics. We advocate that flux datasets need to be used with footprint awareness, especially in research and applications that benchmark against models and data products with explicit spatial information. We propose a simple representativeness index based on our evaluations that can be used as a guide to identify site-periods suitable for specific applications and to provide general guidance for data use.

## 1. Introduction

Global and regional networks of eddy-covariance towers such as those within the FLUXNET and the AmeriFlux provide the largest synthesized *in situ* datasets of energy, water, carbon, and momentum fluxes between Earth’s surface and the atmosphere. By providing a global network of calibration and validation sites, flux networks serve as the cornerstone of the Earth Observing System (EOS) for global terrestrial vegetation monitoring (Running et al., 1999). The large datasets of greenhouse gas and energy surface-atmosphere fluxes are widely used to benchmark Earth system models (Chen et al., 2018; Ricciuto et al., 2018) and flux data products that are based on satellite remote-sensing and machine-learning algorithm upscaling (Tramontana et al., 2016; Verma et al., 2015). While benchmarking models against flux data helps identify key model shortcomings and guide their development, the value of comparisons is greatest when the data are used to understand which processes matter at which spatial and temporal scales. This so-called

“space-time representativeness issue” remains one of the major challenges facing model-data benchmarking (Durden et al., 2020; Hoffman et al., 2017). Adopting the definition in Nappo et al. (1982), representativeness describes the extent to which a set of (flux) measurements taken in a given space-time domain reflect the actual (flux) conditions in a different space-time domain. If focusing on the spatial aspect, the representativeness includes: 1) the network-to-region representativeness (Hargrove et al., 2003), i.e., to what extent do flux measurements taken at a relatively sparse network of tower locations reflect the aggregated flux conditions in a regional or global domain? 2) the point-to-area representativeness (Schmid, 1997), i.e., to what extent do flux measurements taken at a point (tower) location reflect the aggregated conditions over an area that is represented by a model- or satellite-based grid cell? The point-to-area representativeness is of primary interest in the present manuscript. Our primary focus is on evaluating flux data’s representativeness and realizing that similar issues exist in models and other datasets (e.g., satellite data).

While data from eddy-covariance flux towers are well recognized for their rich temporal information (Baldocchi et al., 2001; Katul et al., 2001; Stoy et al., 2009), few eddy-covariance sites are truly homogeneous, and thus, the spatially dynamic nature of these data is often overlooked by modeling and synthesis communities (Giannico et al., 2018; Griebel et al., 2020). Briefly, the extent of the source area contributing to the flux at each observation time—the flux footprint—depends on the wind direction, relative measurement height, underlying vegetation- and ground-surface characteristics, and turbulent state of the atmosphere (Schmid, 2002). When aggregating over time, the flux footprint climatology is essential to identify the spatial extents and temporal dynamics of the areas contributing to the observed fluxes at a tower site (Amiro, 1998). While the footprint concept has gained recognition among the flux community (e.g., Chasmer et al., 2011; Helbig et al., 2016; Wang et al., 2006) and has been evaluated in a few regional-scale studies (Chen et al., 2012; Göckede et al., 2008; Wang et al., 2016), it was often omitted under the assumption of homogeneity at flux tower sites or out of necessity by large-scale syntheses and model-data benchmarking efforts (e.g., Heinsch et al., 2006; Jung et al., 2009; Verma et al., 2015). The main challenges include the lack of footprint information for many sites and difficulty in interpreting and incorporating footprint information into these modeling efforts (i.e., what flux towers observe vs. what models assume). Since most flux tower sites are located in more-or-less heterogeneous landscapes (Griebel et al., 2020; Stoy et al., 2013), the lack of "footprint awareness" leads to unknown biases and uncertainties in modeling and synthesis research (Metzger, 2018).

Some approaches to addressing the spatial mismatch in previous model-data benchmarking research are summarized here. First, Earth system models typically have a model grid cell of around  $10^8 \text{ m}^2$  (Hoffman et al., 2017; Williams et al., 2009) or  $10^6 \text{ m}^2$  in a few regional modeling cases (Buotte et al., 2019; Law et al., 2018), which are often several orders of magnitude larger than the flux footprints of around  $10^3$ – $10^7 \text{ m}^2$ . Scale mismatch remains one of the key challenges and active research topics in model-data benchmarking (Durden et al., 2020; Hoffman et al., 2017; Metzger, 2018; Xu et al., 2020). Model-data benchmarking is often carried out using point scale simulations at individual sites (e.g., Bonan et al., 2012; Chaney et al., 2016; Chen et al., 2018), prescribing with local meteorological forcing data, plant functional type, and other site characteristics that neglect the heterogeneity within the flux footprint. In some cases, model simulations use spatially-explicit gridded forcing data (e.g., satellite-based leaf area index, climate reanalysis data) and assume forcing data from the nearest grid cell (e.g.,  $\sim 10^5$ – $10^9 \text{ m}^2$ ) to represent the conditions at individual sites (Ricciuto et al., 2018; Williams et al., 2020).

Second, remote-sensing data products and models that utilize moderate spatial resolution satellite retrievals (e.g., Moderate Resolution Imaging Spectroradiometer—MODIS) have a grid cell of around  $10^5$ – $10^6 \text{ m}^2$ , which is close to the spatial scale of typical flux footprints. The challenge is to incorporate the temporally dynamic footprints into the modeling frameworks. The moderate resolution can be insufficient for capturing the fine-scale spatial variation, as seen by the flux footprints in some cases (Robinson et al., 2018; Wagle et al., 2020; Yang et al., 2020). Many of the common global, gridded satellite products are themselves aggregated in time and space, e.g., via spatial averaging and compositing over 16-day periods, which further compound the problem of matching flux to remote-sensing data, particularly during periods of disturbance or seasonal transition (Xin et al., 2013). Typically, a fixed set of pixels near the tower locations are selected (Heinsch et al., 2006; Verma et al., 2015; Zhang et al., 2017), assuming the flux conditions in the predefined areas represent the flux conditions observed by the flux footprints. Similar approaches and assumptions are also adopted in studies to extrapolate flux data to spatially-explicit data products (Jung et al., 2019; Xiao et al., 2014; Xiao et al., 2011) and to assess the representativeness of eddy-covariance networks at the regional and global scale (Chu et al., 2017; Villarreal et al., 2018; Villarreal et al.,

2019).

Third, recent advances in fine resolution (e.g.,  $\sim 10^2$ – $10^3 \text{ m}^2$ ) remote-sensing data products (e.g., ECOSTRESS, IKONOS, SPOT, CubeSats, Sentinel, MODIS-Landsat fusion) now enable the flux and remote-sensing data integration at a spatial scale sufficient to resolve the spatial variation as seen by the flux footprints (Anderson et al., 2018; Fisher et al., 2020; Yang et al., 2017). However, a similar approach that selects a fixed set of pixels near the tower locations is still commonly adopted because it lacks footprint information across sites. Alternatively, a few studies adopt a prognostic approach pairing flux data with land surface characteristics at carefully selected pixels based on the prevailing wind direction and estimated source areas (DuBois et al., 2018; Fisher et al., 2020).

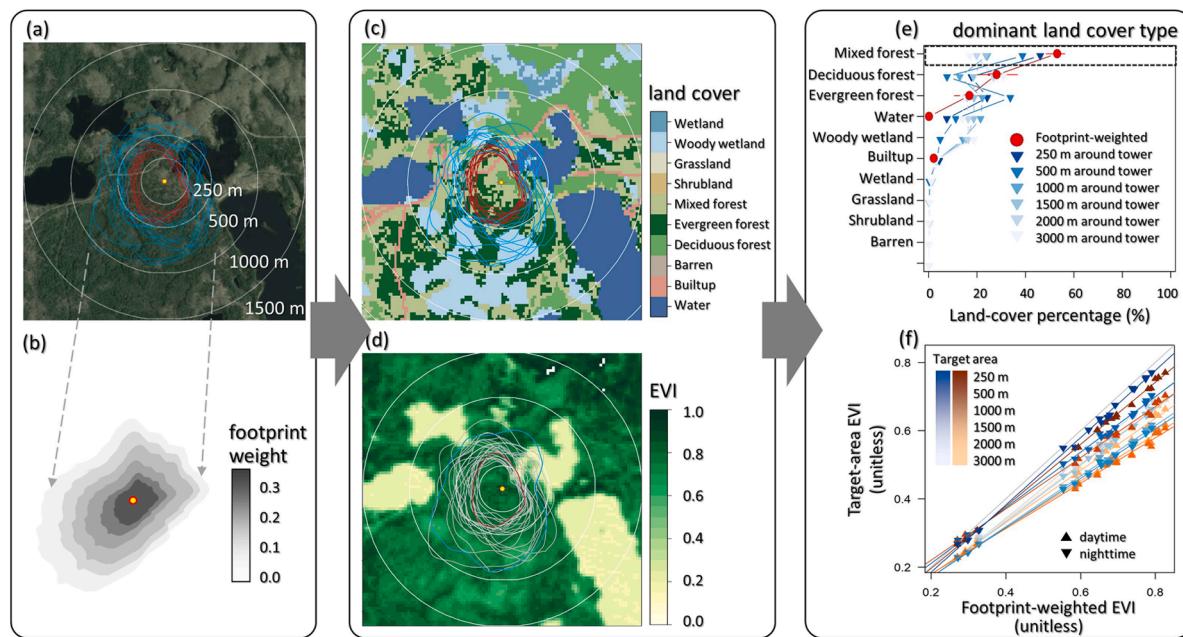
To date, network-wide footprint information is still unavailable or sparse in flux data products, including La Thuile 2007 and FLUX-NET2015 (Papale et al., 2006; Pastorello et al., 2020; Pastorello et al., 2017), which leads to unknown biases and uncertainties in the modeling and synthesis research as discussed above. While several cross-site footprint analyses existed (Chen et al., 2011; Chen et al., 2012; Göckede et al., 2008; Rebmann et al., 2005; Wang et al., 2016), none were carried out extensively across a large number of sites, e.g., FLUX-NET, AmeriFlux. Also, the outcome of the footprint analyses was often not tailored for use by the modeling and synthesis communities. Fortunately, the flux measured at any given time can be traced back to its sources (i.e., spatial locations), providing us with an opportunity to connect the fluxes with land surface characteristics (e.g., land cover type, vegetation characteristics, landform). This study aims to reveal the importance of the network-wide flux footprint information to the modeling and synthesis communities. Our objectives are to 1) calculate and present the variation in the spatial distribution and extent of flux footprints across AmeriFlux sites, 2) evaluate the representativeness of flux footprints to target areas (i.e., areas within 250 m, 500 m, 1000 m, 1500 m, 2000 m, and 3000 m radii around flux towers) that are commonly used in modeling and synthesis studies, and 3) construct simple site-specific representativeness indices that indicate the footprint-to-target-area representativeness and improve the uses and interpretation of flux data by the modeling and synthesis user community.

## 2. Materials and Methods

### 2.1. Overview

To achieve our objectives, we: 1) calculated footprint climatologies for each site, 2) retrieved land surface characteristics in the footprint climatology areas and for a series of predefined target areas at each site, and compared land-surface characteristics between the footprint climatologies and the target areas, and 3) evaluated footprint-to-target-area representativeness as described in Figure 1 and the subsequent sections.

Monthly footprint climatologies were calculated for 214 AmeriFlux sites using a two-dimensional footprint model (Figure 1a–1b, Section 2.2) (Kljun et al., 2015). Two land surface characteristics—one categorical (land cover type) and one continuous (vegetation index)—were retrieved at a fine resolution for each site to match the periods of footprint climatologies (Figure 1c–1d, Section 2.3). These land surface characteristics were used as proxies for representing the spatial variation of land conditions for the fluxes. We defined a series of target areas within 250 m, 500 m, 1000 m, 1500 m, 2000 m, and 3000 m radial distances from the tower at each site. The footprint-to-target-area representativeness was evaluated by comparing the land surface characteristics covered by the footprint climatologies to that in the target areas (Figure 1e–1f, Section 2.4). These target areas were selected based on the spatial resolutions of grid cells/pixels used in previous modeling and synthesis studies. Finally, we proposed qualitative indices for the footprint-to-target-area representativeness for each AmeriFlux site.



**Fig. 1.** Conceptual diagram illustrating the workflow for footprint-to-target-area representativeness analyses using data from US-Syv (2006), as an example, including the calculation of monthly footprint climatology (left panel), the overlay of monthly footprint climatology with land surface characteristics (middle panel), and the comparison of land surface characteristics between the footprints and target areas (right panel). Figures (a), (c), and (d) show the monthly daytime (red) and nighttime (blue) footprint climatologies (80% contour lines) overlaid on (a) a true-color Landsat satellite image, (c) a land cover classification map (e.g., National Land Cover Dataset products (NLCD)), and (d) an enhanced vegetation index (EVI) map (Landsat). Figure (b) illustrates a monthly footprint climatology. Yellow crosshairs in (a)–(d) indicate the tower location. White circles indicate boundaries of selected target areas at 250 m, 500 m, 1000 m, 1500 m radii. Figure (d) is the monthly footprint climatologies that cover the retrieval time of the underlying EVI map, with highlights in red and blue colors while the climatologies from other months are shown in grey colors. Figures (e) and (f) show the results of land cover percentages and EVI calculated from the footprints and target areas, which are then used to evaluate footprint-to-target-area representativeness.

## 2.2. Site selection and footprint calculation for AmeriFlux sites

This study included 214 AmeriFlux sites that had all the required data (i.e., micrometeorological and land surface characteristics) and metadata (e.g., instrument height) (Tables S1 for a site list; Table S2 for site-selection criteria). The selected sites covered 13 International Geosphere-Biosphere Programme (IGBP) vegetation classes (Loveland et al., 1999) and spanned across 13 level-I ecoregions of North America (McMahon et al., 2004; Omernik, 2004). For clarity of visualization, we grouped the IGBP vegetation classes into 6, including evergreen needleleaf forests (59 sites; IGBP: ENF), broadleaf and mixed forests (32; DBF, EBF, MF), savannas and shrublands (28; WSA, SAV, CSH, OSH), grasslands and barren lands (42; GRA, BSV), croplands (34; CRO), and wetlands and snow/ice cover (19; WET, SNO). We grouped the ecoregions into 6, including boreal (42), eastern temperate (43), great plains (39), northwestern mountains and coasts (27), southern desert and semi-arid highlands (29), and Mediterranean California (21). Thirteen sites from the tundra, taiga, Hudson Plain, and tropical wet forests were not included in the ecoregion subgroups but still presented in the pooled results.

Micrometeorological data for each site were obtained from the AmeriFlux BASE data product (<https://ameriflux.lbl.gov/>), accessed in December 2019), including horizontal wind speed, wind direction, friction velocity, Obukhov length, air temperature, humidity, incoming shortwave radiation, and sensible heat flux. Additionally, the standard deviation of cross-wind velocity (V\_SIGMA) was obtained for 106 sites from the AmeriFlux BASE data product, or by contacting the site investigators. Most sites' data had a half-hourly resolution, except the US-MMS, US-Ha1, US-Cop, US-Ne1, US-Ne2, and US-Ne3 sites, which had an hourly resolution. All of the following analyses used the data at their original temporal resolution (i.e., half-hourly, hourly). For sites with short data records (< 3 years), all available years were used. For sites

with long data records, we selected 3–6 years based on data coverage and also the availability of the land surface characteristics (see Table S2 for details of site-year selection and data filtering). Altogether, 712 site-years were used in the study, with the majority of years falling between 2001–2018.

We adopted the Flux Footprint Prediction (FFP) model (<http://footprint.kljun.net/>) of Kljun et al. (2015) for the footprint calculations. The FFP model is a dimensionalized parameterization of a backward Lagrangian stochastic dispersion model (Kljun et al., 2002), which applies to a wide range of boundary layer stratifications and measurement heights (Kljun et al., 2015). Most required variables for the FFP model were obtained from the AmeriFlux dataset as described above, with the following exceptions. Temporal changes in roughness lengths and zero-plane displacement heights were derived following Chu et al. (2018) weekly for cropland, grassland, and wetland sites where seasonal dynamics of canopy heights were expected, and annually for the shrubland, savanna, and forest sites. The planetary boundary layer height was calculated following Kljun et al. (2015), who used the Obukhov length, friction velocity, air temperature, and sensible heat flux in the calculations. For V\_SIGMA, we developed an empirical model using the random forests machine-learning algorithm to predict V\_SIGMA for sites with no data available (Text S1, Figures S1–S2, Table S3). The model was trained, validated, and tested using the 106 sites with available V\_SIGMA data. The model used six predictor variables (i.e., friction velocity, planetary boundary layer height, wind speed, incoming shortwave radiation, Obukhov stability parameter, and IGBP classifications) and showed robust performance against a withheld test dataset ( $R^2 = 0.79$ , mean absolute error (MAE) =  $0.15 \text{ m s}^{-1}$ ) and an independent dataset ( $R^2 = 0.77$ , MAE =  $0.16 \text{ m s}^{-1}$ ) collected using the portable eddy-covariance system by the AmeriFlux Tech Team through 42 site visits (<https://ameriflux.lbl.gov/tech/site-visits/>) (Table S4). Details of the model validation and testing are described in the

Supplement (Text S1, Figures S1–S2, Table S3).

The footprint calculations were carried out for each (half-)hour, generating a two-dimensional gridded map of footprint weights centered on the tower. All available footprint-weight maps in a month were aggregated to daytime and nighttime footprint climatologies (Amiro, 1998) (Figure 1a–1b). Daytime and nighttime were separated by the potential incoming radiation (i.e.,  $> 0 \text{ W m}^{-2}$ ) calculated based on a site's geolocation and time zone. In total, around 14,000 monthly footprint climatologies were generated for all sites combined (Datasets S1–S3). Because the uncertainty of footprint models increases with upwind distance from the receptor (Kljun et al., 2015), we truncated the footprint climatologies at the 80% contour of source weights for all subsequent analyses (Kim et al., 2018). As Schmid (1997) suggested, a point source located outside the 50% contour of source weights must be at least 5–10 times stronger than the point source at the maximum source weight location to achieve a similar sensor response. Our sensitivity tests also showed that the 80% contour cutout selection had a marginal influence on the final results (Text S2, Figures S3–S7, Table S5).

To summarize the footprint climatologies across sites, the following metrics were calculated. The footprint fetch ( $X_{80\_day}$ ,  $X_{80\_night}$  (daytime, nighttime)) was defined by the maximal distance from the tower location to the 80% footprint climatology contour. The footprint area ( $A_{80\_day}$ ,  $A_{80\_night}$ ) was defined by the areas enclosed by the 80% contour. Both footprint fetch and area were averaged from all available monthly footprint climatologies in a site-year. The footprint symmetry index ( $S_{80\_day}$ ,  $S_{80\_night}$ ) was calculated using the following equation:

$$S_{80\_day} = \frac{A_{80\_day}}{\pi \times (X_{80\_day})^2} \quad (1)$$

Equation 1 applies equivalently to  $S_{80\_night}$ . The symmetry index ranges from zero to one, with a value of one indicating a perfectly symmetric (circular) footprint climatology centered around the tower location. The footprint overlap indices included the seasonal overlap index ( $O_{80\_season\_day}$ ,  $O_{80\_season\_night}$ ) and daytime-nighttime overlap index ( $O_{80\_day\_night}$ ) based on the following equations. Both overlap indices range from zero to one, with a value of one indicating perfectly overlapped footprint climatologies either across months (i.e., seasonal overlap) or between daytime and nighttime for each month. The overlap indices were calculated based on all available monthly footprint climatologies in a site-year.

$$O_{80\_season\_day} = \sum_{i=1}^I \left( \prod_{k=1}^K \varphi_{ik}^{day} \right)^{1/k} \quad (2)$$

$$O_{80\_day\_night} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^I (\varphi_{ik}^{day} \times \varphi_{ik}^{night})^{1/2} \quad (3)$$

where  $i$  and  $k$  designate one pixel and one month,  $\varphi_{ik}^{day}$  and  $\varphi_{ik}^{night}$  denote the daytime and nighttime footprint weights at  $i$ -pixel and  $k$ -month, respectively.  $I$  and  $K$  denote the total number of pixels within the footprint climatology and the number of months in a site-year, respectively. As we truncated the footprint climatologies at the 80% contours, the footprint weights were rescaled to sum up to unity within the 80% contours (e.g.,  $\sum_{i=1}^I \varphi_{ik}^{day} = 1$ ). The nighttime overlap index was calculated similarly using Equation 2.

### 2.3. Land surface characteristics

Two types of land surface characteristics—one continuous vegetation index and one categorical land cover type—were used in this study. We chose the enhanced vegetation index (EVI) derived from the Landsat images due to availability. EVI is closely related to vegetation and land cover variations, such as leaf area index (LAI), canopy type, plant

physiognomy, and canopy architecture (Huete et al., 2002). Additionally, EVI is an important land surface product often used in upscaling tower-observed  $\text{CO}_2$  fluxes and as a surrogate of  $\text{CO}_2$  fluxes when the true fluxes (e.g., spatial-explicit) are unavailable (Chen et al., 2011; Fu et al., 2014; Stoy et al., 2013; Xiao et al., 2014). Landsat's 30 m nominal spatial resolution allows the characterization of landscape heterogeneity within the flux footprints. Landsat's 16-day revisit cycle also allowed us to obtain multiple images per year. For each site, we collected all cloud-free (<1% within a 3000 m radius from the tower) Landsat-5/TM and Landsat-8/OLI scenes (USGS Collection Tier 1 atmospherically corrected surface reflectance) that overlapped with the flux measurement periods. Landsat-7 ETM+ data were not considered due to large data gaps caused by the failure of Scan Line Corrector after 31 May 2003. Landsat-5/TM and Landsat-8/OLI provided sufficient data for this study. After the collection of Landsat images, EVI was computed using the following equation (Huete et al., 2002):

$$EVI = 2.5 \times \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + 6 \times \rho_{RED} - 7.5 \times \rho_{BLUE} + 1)} \quad (4)$$

where  $\rho_{NIR}$ ,  $\rho_{RED}$ , and  $\rho_{BLUE}$  are the surface reflectance of near-infrared, red, and blue bands of the Landsat images, respectively. We automated the aforementioned process of image retrieval and EVI calculation in the Google Earth Engine cloud-computing platform. Altogether, around 3,300 images were processed for the 214 AmeriFlux sites.

We chose land cover type as the categorical characteristic because it is commonly used in modeling and upscaling studies (e.g., Fu et al., 2014; Williams et al., 2009; Xiao et al., 2014). The land cover products used in this study include the 2001–2016 United States National Land Cover Dataset products (NLCD; <https://www.mrlc.gov/>) (Yang et al., 2018) and 2010 Land Cover of Canada (<https://open.canada.ca/>) (Latifovic et al., 2017). Both products were derived from Landsat imagery with a spatial resolution of 30 m. For NLCD, we selected the year that overlapped with or approximated the period when the flux data were collected. Because the two land cover data products had slightly different land cover classifications, we merged and consolidated the land cover types into 16 groups (Table S6).

### 2.4. Evaluation of footprint-to-target-area representativeness

The footprint-to-target-area representativeness was evaluated at each site based on the comparisons of EVI and land-cover composition between the footprints and the series of target areas within 250 m, 500 m, 1000 m, 1500 m, 2000 m, and 3000 m radial distances from the tower (Figures 1e–1f, S8–S9).

For land-cover evaluations, we first identified the dominant land-cover type within the footprints, i.e., the land-cover type that had the highest footprint-weighted percentage based on all available monthly footprint climatologies ( $P_{footprint}$ ). Then, the corresponding percentage of this dominant land-cover type in the target areas ( $P_{target}$ ) was calculated. The evaluation of representativeness was based on  $P_{footprint}$ ,  $P_{target}$ , and the Chi-square test between the footprint-weighted and target-area land-cover compositions. For simplicity, we propose a three-level index illustrating the site-level footprint-to-target-area representativeness as follows.

- High representativeness: A site had  $\geq 80\%$  of a certain land-cover type both within the footprint and in a target area, and the land-cover compositions were not significantly different ( $p \geq 0.05$ ) between the footprint and a target area (see an example in Figure S8a–S8b).
- Medium representativeness: A site had  $\geq 50\%$  of a certain land-cover type both within the footprint and in a target area, and the land-cover compositions were not significantly different ( $p \geq 0.05$ ) between the footprint and a target area, if not meeting the criteria of high representativeness (Figure S8c–S8d).

- Low representativeness: A site had  $< 50\%$  of a certain dominant land-cover in the footprint or target area, or the land-cover compositions were significantly different ( $p < 0.05$ ) between the footprint and a target area (Figure S8e–S8f).

The criteria of 50% and 80% were chosen following Göckede et al. (2008). All statistics of  $P_{footprint}$ ,  $P_{target}$ , and Chi-square test are provided in the supplement (Dataset S4). In the preliminary tests, we found that 34 sites had incorrect land-cover classifications from the land-cover products (e.g., NLCD) as compared with sites' metadata (e.g., IGBP) (details in Text S3). Thus, the site-specific representativeness of land-cover composition was quantified at only 180 sites.

For EVI evaluations, we first matched the available Landsat EVI images with the corresponding month's footprint climatologies. At each available month, we calculated the target-area mean EVI ( $EVI_{target}$ ), footprint-weighted EVI ( $EVI_{footprint}$ ), and the sensor location bias ( $\Delta$ ) after Schmid and Lloyd (1999):

$$EVI_{footprint} = \sum_{j=1}^J (\varphi_j \times EVI_j) \quad (5)$$

$$\Delta = \frac{EVI_{footprint} - EVI_{target}}{EVI_{target}} \quad (6)$$

where  $j$  designates a pixel,  $J$  denotes the total number of pixels within the footprint climatology,  $\varphi_j$  and  $EVI_j$  denote the footprint weight and EVI at pixel  $j$ , respectively. The sensor location bias, which was adopted in previous studies (Chen et al., 2011; Kim et al., 2006; Kim et al., 2018), represented the time-explicit footprint-to-target-area bias of EVI in each available month. Additionally, we quantified the site-level footprint-to-target-area representativeness using linear regression (model II) between all available footprint-weighted and target-area EVI at a site:

$$EVI_{target} \sim \beta_0 + \beta_1 \times EVI_{footprint} \quad (7)$$

where  $\beta_0$  and  $\beta_1$  denote the intercept and slope of regression, respectively. Similarly, we propose a three-level representativeness index based on the regression results as follows.

- High representativeness: A site had linear regression with  $R^2 \geq 0.8$ , a slope of  $1.0 \pm 0.1$ , and an intercept of  $0.0 \pm 0.1$  between the footprint-weighted and target-area EVI (see an example in Figures S9a–S9b).
- Medium representativeness: A site had linear regression with  $R^2 \geq 0.6$  and  $p < 0.05$ , if not meeting the criteria of high representativeness (see an example in Figures S9c–S9d).
- Low representativeness: A site had linear regression with  $R^2 < 0.6$  or  $p \geq 0.05$  (Figures S9e–S9f).

Different thresholds (e.g., 5%, 10%) of sensor location bias have been used to justify the representativeness in the previous studies (Chen et al., 2011; Kim et al., 2006; Wang et al., 2016). Here, we adopted a similar 10% threshold for both the sensor location bias in each available month (i.e.,  $|\Delta| \leq 10\%$ ) and site-level regression (i.e.,  $-0.1 \leq \beta_0 \leq 0.1$ ,  $0.9 \leq \beta_1 \leq 1.1$ ). All statistics of  $EVI_{footprint}$ ,  $EVI_{target}$ ,  $\Delta$ , and regression are provided in the supplement (Datasets S5–S6). While all 214 sites were used in the pooled analyses, site-specific regressions were carried out for only 166 sites with at least six matches between the monthly footprint climatologies and Landsat EVI. The median of available matches was 13 per site for the 166 sites.

Unless specified, all data processing and statistical analyses were conducted using R (R Core Team, 2019). Specifically, the R-version of the FFP model was downloaded through <http://footprint.kljun.net/>. The h2o package was used to interface with H2O (<https://www.h2o.ai/>), an open-source cloud platform for distributed and scalable machine learning, to carry out the random forests model training and validation.

The model II linear regression (lmodel2 package) was adopted for the comparison of EVI and evaluation of V\_SIGMA prediction (Legendre, 2014). The ggmap and rasterVis packages were used in generating the footprint maps (Kahle and Wickham, 2013; Lamigueiro and Hijmans, 2018).

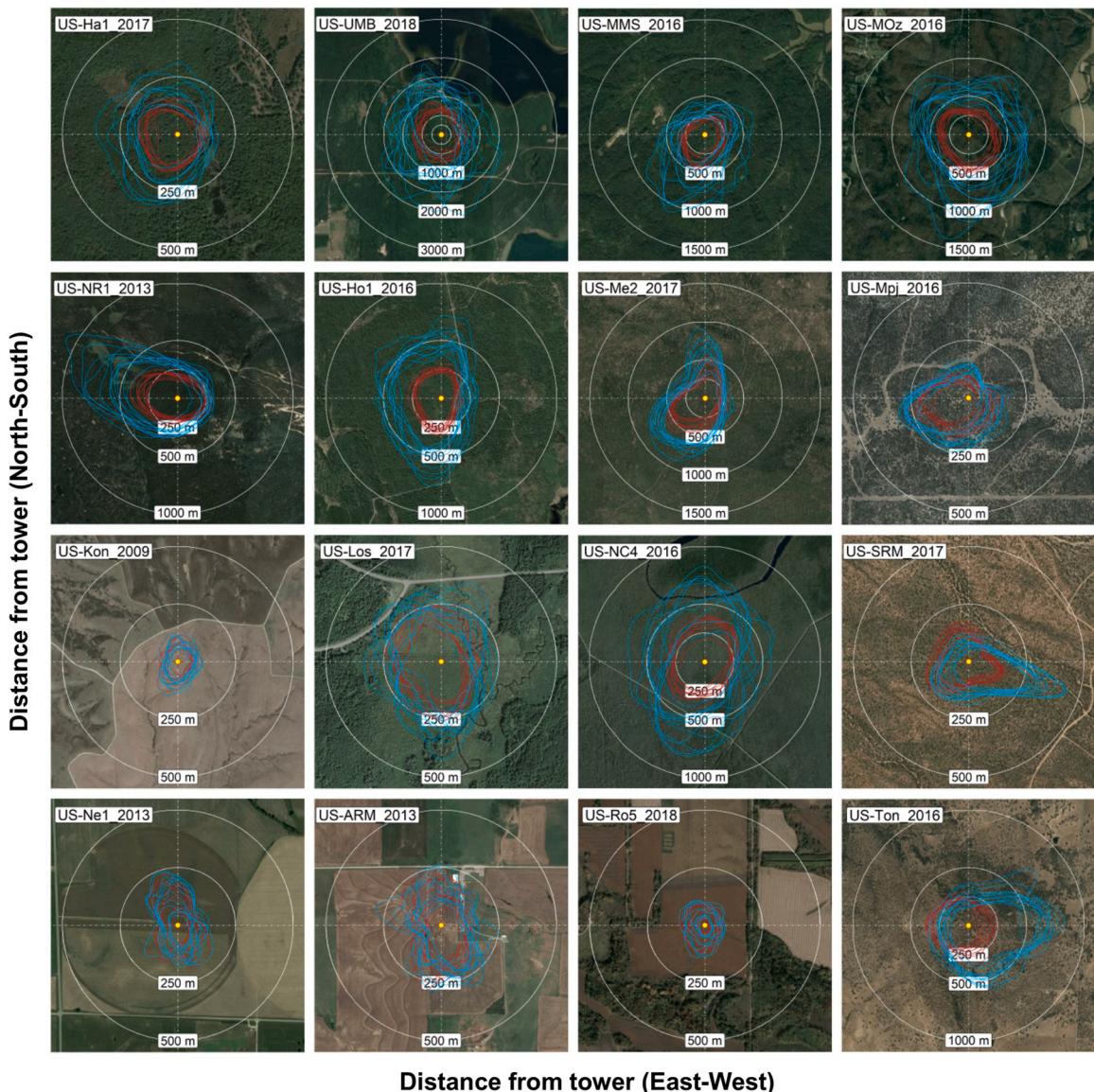
## 3. Results

### 3.1. Footprint climatology across AmeriFlux sites

Across AmeriFlux sites, the spatial extent of flux footprint climatologies spanned several orders of magnitude (Figure 2, S10–S17). The footprint climatology extents, defined by the maximal distance from the tower to the 80% contour of the monthly footprint climatologies, ranged mostly from 100 m to 450 m (25th and 75th percentiles, Figure 3a). The footprint climatology areas, defined by the 80% contour of the monthly footprint climatologies, covered mostly around 13,000 to 260,000 m<sup>2</sup> in area (25th and 75th percentiles, Figure 3b). In almost all studied site-years ( $> 95\%$ ), the nighttime footprint climatologies extended farther ( $\sim 45\%$  on average) and covered a larger area ( $\sim 90\%$  on average) than the daytime footprint climatologies. In rare cases, the daytime footprint climatologies may cover a slightly larger area than nighttime due to the distinct and variable wind directions between the daytime and nighttime (Text S4, Figure S18–S19). As the daytime and nighttime footprint climatologies largely overlapped in the areas with higher weights, around 93% of the site-years had  $O_{80\_daynight}$  larger than 0.8 (Figure 3d, Text S2). Additionally, footprint climatology, which aggregated footprints from many time steps from different wind directions, smoothed out the variable and potentially extreme footprint conditions as seen by the half-hourly footprints (Figure S19). That also explained the relatively high  $O_{80\_daynight}$  at most studied sites. While most sites had a relatively symmetric footprint climatology centered around the tower locations, some sites were asymmetric (e.g., US-NR1 ( $S_{80\_night} = 0.31$ ), US-SRM ( $S_{80\_night} = 0.18$ ) in Figure 2), as a consequence of the unimodal or bimodal prevailing wind directions. Using an arbitrary criterion of 0.30 for the symmetry index, around 7% and 15% of the studied site-years showed relatively asymmetric footprint climatologies for daytime and nighttime, respectively (Figure 3c). Noticeable monthly variability was found at  $\sim 32\text{--}44\%$  of the studied site-years (i.e.,  $O_{80\_season\_day}$ ,  $O_{80\_season\_night} < 0.8$ , Figure 3d). Specifically, cropland, grassland, and wetland sites that experienced large changes of canopy heights throughout the growing season showed relatively pronounced monthly variations in footprint climatologies (e.g., US-Ne1, US-ARM, US-Ro5 in Figure 2).

### 3.2. Evaluation of land-cover type

While most AmeriFlux sites (83%–84%, i.e., daytime, nighttime) had relatively homogeneous land cover within the flux footprint climatologies, only 25%–39% of sites had similar land-cover compositions in the areas extending kilometers from the towers. Around 84% and 83% of AmeriFlux sites had one dominant land-cover type, as seen by the daytime and nighttime footprint climatologies, respectively (i.e., the dominant land-cover type accounted for  $\geq 80\%$  footprint-weighted land-cover percentages (Figure 4)). On the other hand, only around 64% of the sites had the same land-cover type also dominating (i.e.,  $\geq 80\%$ ) the 250 m target areas. The percentages of sites decreased rapidly as the target area extended farther from the towers. With the 1000 m and 3000 m target areas, only 39% and 25% of sites had the same dominant land-cover type as seen by the flux footprint climatologies. The median difference of the dominant land-cover percentages between the footprint and target area was  $\sim 6\%$  and  $\sim 20\%$  ( $P_{footprint} > P_{target}$ ) at the 1000 m and 3000 m target areas, respectively. These results suggest that most AmeriFlux sites represent a specific land-cover type at each site, but they may not represent the dominant land-cover type at a larger spatial extent (e.g., 1000 m, 3000 m) around the tower.



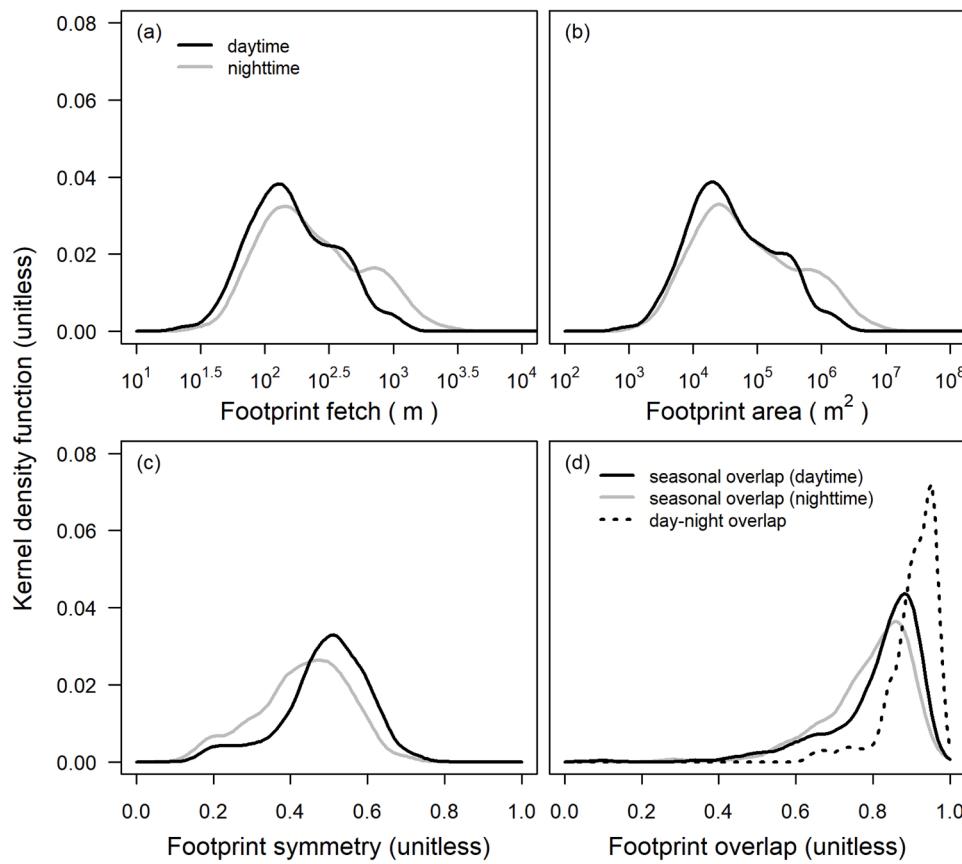
**Fig. 2.** Maps of monthly footprint climatologies at 16 AmeriFlux Core sites. In each panel, a true-color satellite image (accessed through Google Map) centered on a flux tower (yellow crosshairs) was overlaid with monthly footprint climatologies (i.e., 80% contour, red: daytime, blue: nighttime) from a selected year. White circles denote the distance from the tower. For better visualization, each map was truncated at different target areas. The site ID and selected year are indicated in the inserted label. Similar maps for other AmeriFlux sites are provided in the Supplement (Figures S10–S17).

For site-specific evaluations of land-cover composition, around 64% and 65% of sites showed high representativeness for the 250 m target area based on daytime and nighttime footprints (Figures 5a, S20a), and the percentage dropped to 21%–27% and 19%–28% as the target areas extended beyond 1000 m from the towers. The percentage of high-representativeness sites also varied substantially among IGBP and ecoregion groups (Figures 5, S20). Cropland sites, mostly located in highly managed agricultural landscapes, had 90%–50% of high-representativeness sites for 250–3000 m target areas (Figure 5e, S20e). On the other hand, only half of the wetland sites showed high representativeness even for a 250 m target area (Figures 5g, S20g, examples in Figures S17 (US-ORv, US-Srr, US-StJ)). The percentage of high-representativeness sites dropped to ~20% when extending the target areas beyond 500 m from the towers. For savanna and grassland sites, the percentage of high-representativeness sites was around 50%–60% within the 500 m target area. It gradually decreased to about 16%–33% as the target areas extended beyond 1000 m from the towers. Except for cropland, all ecosystem types had 4%–18% of low-

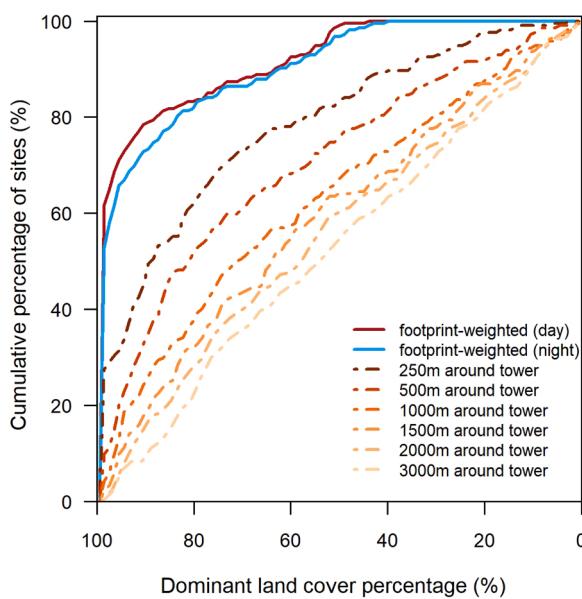
representativeness sites as the target areas extended beyond 1000 m from the towers. Among the ecoregions, southwestern desert/semiarid and Mediterranean California regions had 40%–82% of high-representativeness sites across all target areas (Figures 5l, 5m, S20l, S20m). Eastern temperate and Boreal regions had only 2%–16% of high-representativeness sites when the target areas extended beyond 500 m around the tower.

### 3.3. Evaluation of Enhanced Vegetation Index

Systematic EVI biases (4%–20%) were found between the flux footprint climatologies and the target areas using a fixed area across all available site-years (Figure 6, Table 1). As indicated by the regression slopes, the systematic biases were ~4% and ~9% when using fixed 250 m and 500 m target areas across all site-years (Table 1), respectively. Bias increased as the target area extended farther from the tower, reaching around 20% with a 3000 m target area. At all target areas, footprint-weighted EVI was systematically higher than target-area EVI,



**Fig. 3.** Kernel density functions of the footprint climatology across AmeriFlux sites. Figure (a) shows the distribution of the maximal distance (i.e., fetch ( $X_{80\_day}$ ,  $X_{80\_night}$ )) from the tower location to the 80% footprint climatology contour. Figure (b) shows the distribution of the area within the 80% footprint climatology contour ( $A_{80\_day}$ ,  $A_{80\_night}$ ). Figure (c) shows the distribution of the footprint symmetry index ( $S_{80\_day}$ ,  $S_{80\_night}$ ). Figure (d) shows the distribution of the footprint overlap index for seasonal ( $O_{80\_season\_day}$ ,  $O_{80\_season\_night}$ , solid lines) and daytime-nighttime overlaps ( $O_{80\_daynight}$ , dashed line). When calculating these statistics, a data point represents an averaged value from all available monthly footprint climatologies in a studied site-year. The kernel density function was then obtained using data points from all available site-years. The black and grey lines show the distributions based on daytime and nighttime footprint climatologies, respectively. Both fetches and areas are logarithmic transformed (x-axis) in Figures (a) and (b).



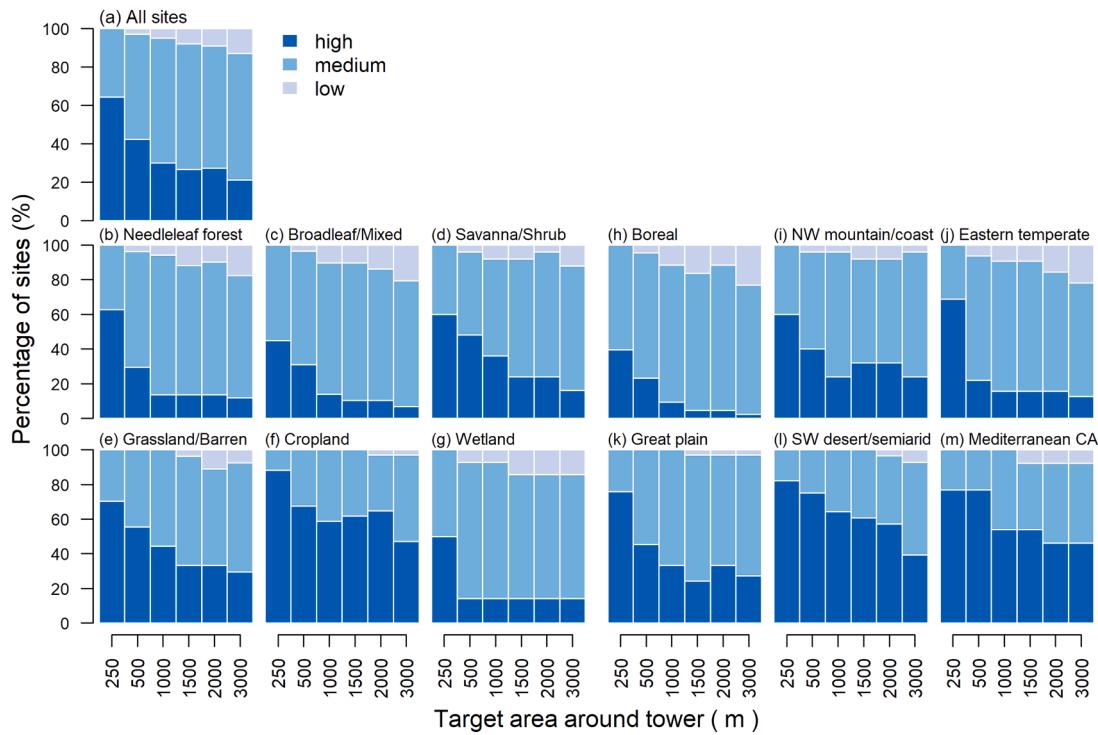
**Fig. 4.** Cumulative distributions of the percentage of dominant land-cover type in the footprints and target areas across AmeriFlux sites. The red and blue lines are based on daytime and nighttime footprint climatologies, respectively. The orange dash-dotted lines show results from a series of target areas (i.e., 250–3000 m radii centered around the tower location) as indicated by the color intensity (from dark to light, indicating an increasing distance from the tower).

suggesting that the flux footprints mostly covered areas with higher EVI than the surroundings. Only marginal differences ( $\sim 1\%$  in regression slopes) were found between results using daytime and nighttime

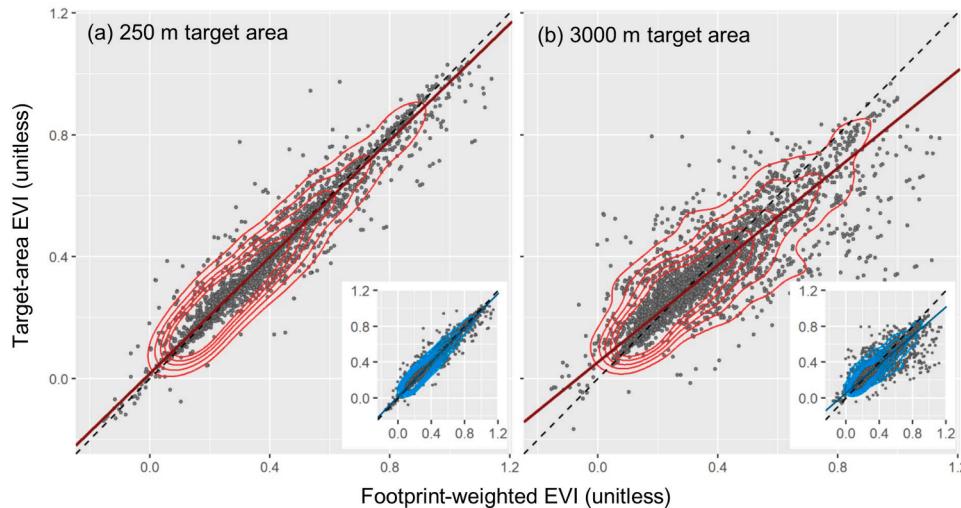
footprint climatologies (insets in Figure 6, Table 1). In addition to the larger systematic biases, the comparisons of footprint-weighted and target-area EVI were more variable (i.e., they exhibited case-by-case, month-by-month variation) with the target areas extending farther from the towers. The  $R^2$  was 0.94 with a 250 m target area and dropped to 0.71 with a 3000 m target area. The results suggest that no single fixed target area can represent the flux footprint climatologies across sites without introducing biases. However, the 250 m target area was less biased in representing the footprint-weighted EVI when pooling all site-years altogether.

The time-explicit sensor location biases ( $\Delta$ ) showed similar dependency on the extents of the target areas. Around 73% and 75% of the monthly sensor location biases were within the  $\pm 10\%$  threshold when using a 250 m target area for daytime and nighttime, respectively (Figure 7). The percentages within the  $\pm 10\%$  threshold decreased as the target areas extended farther from the tower, reaching 42% and 43% for daytime and nighttime at a 3000 m target area. Separating by seasons, the summer months (June–August) had slightly higher percentages within the  $\pm 10\%$  threshold, which were, respectively, 75% and 77% for daytime and nighttime at a 250 m target area and 45% and 46% at a 3000 m target area (Figures S21e–S21f). In contrast, the winter months (December–February) had lower percentages of 70% and 68% for daytime and nighttime at a 250 m target area and 33% and 34% at a 3000 m target area, respectively (Figures S21a–S21b).

For site-specific regressions using EVI, 53% and 54% of sites showed high representativeness for a 250 m target area based on daytime and nighttime footprints (Figure 8a, S22a), and the percentage dropped to 11%–16% and 11%–16% as the target areas extended beyond 1000 m from the towers. In all IGBP and ecoregion groups, the percentages of high-representativeness sites generally decreased as the target areas extended farther away from the tower. In addition, the percentage of high-representativeness sites varied substantially among the ecosystem types and ecoregions (Figures 8, S22). In sum, broadleaf and mixed



**Fig. 5.** The footprint-to-target-area representativeness of the land-cover composition. Figure (a) shows the pooled results from all study sites. Figures (b)–(g) show results by ecosystems type, including the (b) needleleaf forest (i.e., IGBP: ENF), (c) mixed and broadleaf forest (MF, DBF, EBF), (d) savanna and shrubland (WSA, SAV, CSH, OSH), (e) grassland and barren land (GRA, BSV), (f) cropland (CRO), and (g) wetland (WET, SNO). Figures (h)–(m) show results by ecoregion, including the (h) boreal, (i) northwestern mountain and coast, (j) eastern temperate, (k) great plains, (l) southwestern desert and semiarid, and (m) Mediterranean California. The color intensity indicates the representativeness levels (from dark to light, indicating high to low representativeness). The results based on daytime footprints are shown here, while nighttime results are provided in the supplement (Figure S20).



forest sites had the highest percentages of high-representativeness sites across all target areas (Figures 8b, S22b), from ~77% with a 250 m target area to 27–32% when the target areas extended beyond 1000 m from the towers. Savanna and shrubland sites also had relatively high percentages of high- and medium-representativeness sites across all target areas. Cropland sites, while having ~50% of high-representativeness sites with a 250 m target area, decreased drastically as the target areas extended beyond hundreds of meters from the towers. Croplands, needleleaf forests, grasslands, and wetlands had around 20%–40% of low-representativeness sites as the target areas extended beyond 1000 m from the towers. Among the ecoregions, the

great plains had relatively higher percentages of high-representativeness sites across all target areas (Figures 8k, S22k), from ~53% with a 250 m target area to 21–29% when the target areas extended beyond hundreds of meters from the towers. Northwestern mountain/coast, eastern temperate, southwestern desert/semiarid, and Mediterranean California ecoregions had around 20%–67% of low-representativeness sites with a target area extending beyond 1000 m from the towers.

**Table 1**

Pooled results of the comparison of footprint-weighted and target-area EVI for all available site-months ( $N = 3307$ ). Linear regression model:  $\text{EVI}_{\text{target}} \sim \beta_0 + \beta_1 \times \text{EVI}_{\text{footprint}}$ . RMSE: root mean square error. MAE: Mean absolute error. Numbers in square brackets indicate 95% confidence intervals. See [Figure 6](#) for the results from the 250 m and 3000 m target areas.

Target area	Intercept ( $\beta_0$ )	Slope ( $\beta_1$ )	$R^2$	RMSE	MAE
<i>Daytime footprint</i>					
250 m	0.02 [0.01, 0.02]	0.96 [0.95, 0.97]	0.94	0.05	0.03
500 m	0.03 [0.03, 0.04]	0.91 [0.90, 0.92]	0.88	0.07	0.05
1000 m	0.04 [0.04, 0.05]	0.87 [0.86, 0.89]	0.79	0.09	0.06
1500 m	0.05 [0.04, 0.05]	0.85 [0.83, 0.86]	0.74	0.09	0.07
2000 m	0.05 [0.05, 0.06]	0.82 [0.81, 0.84]	0.72	0.09	0.07
3000 m	0.06 [0.05, 0.06]	0.80 [0.78, 0.81]	0.71	0.09	0.07
<i>Nighttime footprint</i>					
250 m	0.02 [0.01, 0.02]	0.96 [0.96, 0.97]	0.94	0.05	0.03
500 m	0.03 [0.03, 0.03]	0.92 [0.91, 0.93]	0.88	0.07	0.04
1000 m	0.04 [0.04, 0.05]	0.88 [0.87, 0.89]	0.79	0.09	0.06
1500 m	0.05 [0.04, 0.05]	0.85 [0.84, 0.86]	0.75	0.09	0.07
2000 m	0.05 [0.04, 0.06]	0.83 [0.81, 0.84]	0.73	0.09	0.07
3000 m	0.05 [0.05, 0.06]	0.80 [0.79, 0.82]	0.71	0.09	0.07

### 3.4. Example Cases

Three typical cases are discussed based on the representativeness for both the land-cover composition and EVI (Figures S23–S26). First, an example demonstrates that a site's flux footprints are representative of most of the target areas. The US-MOz site—a deciduous forest located in a forest-dominated landscape—showed similar evaluation results for both the land-cover composition and EVI at most target areas (Figure S23). The site was classified as high representativeness for the EVI at all target areas and land-cover composition up to a 1500 m target area. At 2000 m and 3000 m target areas, the site was classified as medium representativeness for land-cover composition due to a slightly higher percentage of cropland coverage.

Second, an example demonstrates that a site's flux footprints represent only limited extents from the tower. The US-Vcp site—an evergreen forest located within a forest-shrub-grassland landscape—showed a gradually declining representativeness as the target area extended farther from the tower (Figure S24). The site was classified as high representativeness for both EVI and land-cover composition at a 250 m target area. Extending farther from the tower, the target areas covered more shrublands and grasslands. Consequently, the site was classified as low and medium representativeness for EVI and land-cover composition when the target area extended beyond a 1000 m radius around the

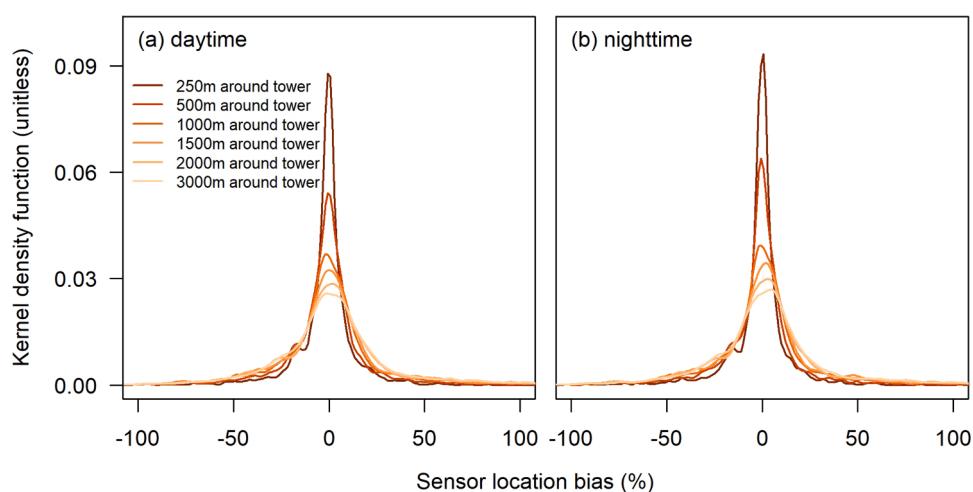
tower. Both US-MOz and US-Vcp cases show similar evaluation results between the land-cover composition and EVI.

Third, contrasting representativeness between land-cover composition and EVI were found in several cropland sites, attributed to the variety of crop types that were not differentiated by the land-cover products. For example, the US-Ro6 site—a cropland located in an agricultural landscape dominated with corn/soybean rotation—was planted with wheat, clover, and corn in 2017–2018 (Figure S25). The site was planted with spring wheat in April 2017. Kura clover was planted after the harvest of wheat in August. Corn was planted into the clover in the spring of 2018. While the site was classified as high/medium representativeness for the land-cover composition across all target areas, the site was classified as low representativeness for the EVI in all except the 250 m target area, as a consequence of the distinct crop types and phenology between the footprints and nearby fields. On the other hand, the US-Bi2 site—a corn cropland located in a corn-dominant agricultural landscape—showed medium to high representativeness for both land-cover composition and EVI at all target areas (Figure S26).

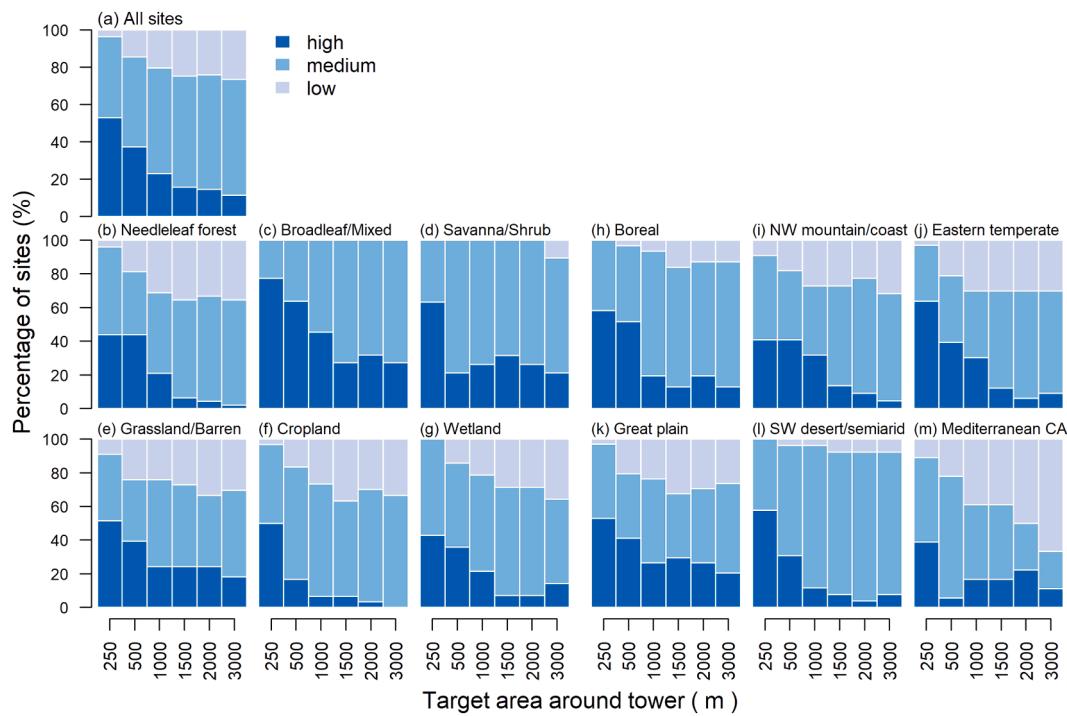
## 4. Discussion

### 4.1. Implication of footprint-to-target-area representativeness

We advise the modeling and synthesis communities to be "footprint-aware" when using the large-scale flux datasets (e.g., FLUXNET2015, AmeriFlux BASE data products), especially in research such as point-scale simulations and spatially-explicit land surface models, remote-sensing-based models, and machine-learning upscaling studies. Our study highlights the importance of considering spatiotemporal dynamics of flux footprints, particularly within heterogeneous landscapes, when using flux data in model-data benchmarking. Across AmeriFlux sites, we found several orders of magnitude difference in the extents and areas of the footprint climatologies. As most flux tower sites were located in more-or-less heterogeneous landscapes (e.g., land cover, EVI) (Griebel et al., 2020; Reuss-Schmidt et al., 2019; Stoy et al., 2013), no single fixed-extent target area can represent the land surface characteristics as seen by the flux footprints universally across sites without introducing biases. These biases' sign and magnitude were site-specific and varied depending on the extents of target areas and the land surface characteristics. In general, the 250 m target area tended to be less biased in representing the land surface characteristics as covered by the flux footprint climatologies. Yet, around 36% and 47% of sites still differed in the land-cover composition and EVI between the footprints and a 250



**Fig. 7.** Kernel density functions of the sensor location bias ( $\Delta$ ) of EVI for all available site-months ( $N = 3307$ ) from (a) daytime and (b) nighttime footprints. The orange lines show results from a series of target areas (i.e., 250–3000 m radii around the tower location) as indicated by the color intensity (from dark to light, indicating an increasing distance from the tower). A similar figure broken down by season is provided in Figure S21.



**Fig. 8.** The footprint-to-target-area representativeness of the EVI. Figure (a) shows the pooled results from all study sites. Figures (b)–(g) show results by ecosystems type, including the (b) needleleaf forest (i.e., IGBP: ENF), (c) mixed and broadleaf forest (MF, DBF, EBF), (d) savanna and shrubland (WSA, SAV, CSH, OSH), (e) grassland and barren land (GRA, BSV), (f) cropland (CRO), and (g) wetland (WET, SNO). Figures (h)–(m) show results by ecoregion, including the (h) boreal, (i) northwestern mountain and coast, (j) eastern temperate, (k) great plains, (l) southwestern desert and semiarid, and (m) Mediterranean California. The color intensity indicates the representativeness levels (from dark to light, indicating high to low representativeness). The results based on daytime footprints are shown here, while the nighttime results are provided in the supplement (Figure S22).

m target area, respectively. With limited footprint information across sites, previous studies often assumed flux data collected at the tower locations adequately represented the conditions at a certain set of fixed areas surrounding the towers (e.g.,  $1 \times 1$  or  $3 \times 3 \text{ km}^2$ ) (Tramontana et al., 2016; Verma et al., 2015; Xiao et al., 2014), or from a specific plant functional type (Chen et al., 2018; Williams et al., 2009). Such assumptions may not hold for many sites and would need to be revisited. We discuss potential approaches to being "footprint aware" in the following sections.

We advocate that the representativeness indices can be used to identify and select sites suitable for specific applications considering the spatial scales of interest and relevant land surface characteristics—a subsetting approach. While simple and straightforward, the subsetting approach utilizes only a portion of available site-data. Also, the criteria of the representativeness levels may not be suitable for certain research and applications. Alternatively, the representativeness indices can be used to provide guidance of data usage for sites with different representativeness levels—a hybrid approach. For sites classified as "high-representativeness" for a given target area, the flux data may be used without additional detailed footprint analyses (after considering the limitations discussed below). Sites classified as "medium-representativeness" for a given target area were often located in a relatively heterogeneous or patchy landscape, and hence a prognostic method, such as pairing flux data with land surface characteristics at carefully selected pixels (i.e., a spatially-explicit method; DuBois et al., 2018; Fisher et al., 2020; Verma et al., 2015; Zhang et al., 2017), might be sufficient. Our monthly footprint climatologies (Figures 2, S10–S17, Datasets S2–S3) could be used to define the areas. Alternatively, a more sophisticated method that incorporates the subgrid variability in model simulations would be necessary to adequately link the observations with model outputs (Balocchi et al., 2005; Bonan et al., 2012). Sites classified as "low-representativeness" for a given target area were typically located in a heterogeneous or patchy landscape, and additionally, were often

characterized by time-varying footprint climatologies and/or distinct vegetation phenology within the landscape (e.g., different crop types in adjacent fields, a mix of different plant functional types, vegetated and non-vegetated cover). We advise that site-specific footprint analysis is needed to use and interpret flux data at these sites. The footprint quality flags recently adopted by AmeriFlux and Integrated Carbon Observation System (ICOS: <https://www.icos-cp.eu/>) or the time-explicit footprint weight maps provided by National Ecological Observatory Network (NEON: <https://www.neonscience.org/>) can be used to identify and filter out periods when flux footprints covered the less desired source areas. Several site-level studies also demonstrated the potential benefits of utilizing detailed footprint information in decomposing flux data signals in a patchy and heterogeneous site (Helbig et al., 2016; Tuovinen et al., 2019; Wang et al., 2006). Alternatively, a spatiotemporally-explicit method that incorporates both the temporal dynamics of footprints and the spatial variations of land surface characteristics is highly recommended (Chen et al., 2010; Fu et al., 2014; Ran et al., 2016; Xu et al., 2017a; Xu et al., 2017b).

Several cross-site footprint analyses were carried out at the network scales previously (Chen et al., 2011; Chen et al., 2012; Göckede et al., 2008; Rebmann et al., 2005; Wang et al., 2016). While incorporating certain aspects or metrics adopted in those previous studies, our approach is tailored to the modeling and synthesis data users who intend to use flux data for many sites. We demonstrate our approach's feasibility to be implemented "extensively" across a large number of sites and propose a simple representativeness index that could be used for site selection and data use guidance. Such an "extensive" approach can help subset data that better fulfill the space-time assumptions underlying a particular application. To address the space-time mismatches between flux observation and model simulation in more detail and utilize (rather than subset) the full set of available sites, we recommend a hierarchy of multiple approaches where feasible (Durden et al., 2020). For example, Griebel et al. (2020) proposed a practical approach to account for

contributions of spatial sampling variations due to varying wind directions over heterogeneous surfaces for annual budgets of carbon fluxes. While the approach does not calculate footprints explicitly, it had relatively few implementation requirements and was successfully applied to 154 sites in the FLUXNET2015 dataset. Alternatively, a few studies proposed relatively "intensive" approaches to fully incorporate the footprint dynamics in footprint-to-target-area upscaling frameworks (Fu et al., 2014; Metzger et al., 2013; Ran et al., 2016; Reuss-Schmidt et al., 2019; Xu et al., 2017b). These approaches show merit for providing spatiotemporally-explicit flux data products that could be benchmarked against model predictions at designated grid cells. One such development even shows promise to close the surface energy imbalance frequently observed at flux towers (Metzger, 2018; Xu et al., 2020), which to date cripples data synthesis and model-data fusion with another pervasive bias (Cui and Chui, 2019; Mauder et al., 2020; Stoy et al., 2013). However, the data requirements for implementing these approaches across sites are comparatively high. These include the availability of fine-resolution (spatially and temporally) land surface characteristics and/or reprocessing the high-frequency eddy-covariance data. Our representativeness results could serve as a guide to identify and prioritize the sites where the intensive approach is deemed necessary (Durden et al., 2020).

#### 4.2. Limitations and uncertainties

The FFP model—an analytical parameterization of a Lagrangian stochastic dispersion model—is robust in performance (Heidbach et al., 2017; Nicolini et al., 2017) and flexible enough to be applied across many sites over long periods. However, the FFP model, as well as most footprint models, has its assumptions and limitations. These include the assumption of steady-state conditions, a horizontally homogeneous turbulence field, no vertical advection, and the applicability of the Monin-Obukhov similarity theory (Rannik et al., 2012; Kljun et al., 2015). While we recognize these assumptions and carefully filter out less ideal sites and/or data periods (Table S2), it is challenging to systematically evaluate these assumptions (especially the last two) across AmeriFlux sites. We briefly discuss the potential uncertainties and implications here. With flux data being carefully quality-controlled/filtered by the site investigators and a large number of (half-)hourly footprints aggregated for monthly footprint climatologies, we argue that the influences of non-steady states should be marginal on the final results. The unmet assumption of homogeneity at certain sites requires caution when interpreting the evaluation results. Similar to previous footprint-based evaluation studies (Chen et al., 2012; Göckede et al., 2008), our evaluation of footprint representativeness is inherently influenced by landscape heterogeneity because it violates the homogeneity assumption of most footprint models. The sites classified as "low-representativeness" might be prone to additional uncertainties resulting from an inhomogeneous turbulence field (Heidbach et al., 2017). Besides, ~14% of the studies sites were located in an undulated terrain, a strong slope (>10%), or a hilltop, and ~28% of the studied sites had the eddy-covariance measurement height within the roughness sublayer, assuming it extends to 1.5 times the aerodynamic canopy height (Chu et al., 2018). The possible occurrence of advection and violation of the Monin-Obukhov similarity theory may complicate the footprint calculations. We suggest using caution when interpreting results from these sites (listed in Dataset S4–S5). Further site-specific footprint analyses are highly recommended for these cases. It should be noted that the footprint models typically assume flux sources/sinks are located dominantly in the plant canopy. Such an assumption implies that our footprint evaluations are less applicable to CH<sub>4</sub> and nighttime CO<sub>2</sub> fluxes, especially at tall forest sites.

Additional uncertainties in footprint calculations could result from the selection of footprint models or uncertainties in the input variables/parameters. First, we recognized that there are other types of footprint models such as the analytical models (e.g., Hsieh et al., 2000; Kormann

and Meixner, 2001; Schmid, 2002; Schuepp et al., 1990) and Lagrangian stochastic models (e.g., Leclerc and Thurtell, 1990; Thomson, 1987). Our preliminary tests found that the extents and areas of footprint climatologies could vary substantially using a different footprint model (e.g., Kormann and Meixner, 2001 (KM01 model)) (Text S2, Figure S3). Yet, as the areas with higher footprint weights were largely overlapped between the FFP and KM01 models, we found only marginal differences in the results of footprint-to-target-area representativeness, footprint-weighted land-cover percentage and EVI (Text S2, Figures S4–S7). Footprint model benchmarking and validation is an active research area beyond the scope of this study, and we refer interested readers to the relevant studies (Arriga et al., 2017; Dumortier et al., 2019; Heidbach et al., 2017; Kumari et al., 2020; Nicolini et al., 2017). We also advocate that future research should aim to better quantify and constrain the uncertainty of footprint models, as ultimately, all footprint models are "models" and have their limitations and uncertainties. Second, we parameterized the roughness lengths and zero-plane displacement heights following Chu et al. (2018). This allowed us to track canopy dynamics (e.g., seasonal dynamics for croplands, Figures S25–S26) and evaluate many sites where required metadata (e.g., canopy structure) are not always available. Yet, the approach has its assumptions and limitations (Chu et al., 2018). Potential uncertainties or biases could result from the violation of the Monin-Obukhov similarity theory, an aerodynamically inhomogeneous turbulence field, or any unaccounted changes in canopy structures. Again, caution should be exercised when interpreting results at sites with complex terrain and relatively low eddy-covariance measurement heights because those sites likely violate the assumptions mentioned above. Third, the use of predicted V\_SIGMA could introduce additional uncertainties in the footprint calculations, especially in the lateral dimension of the footprint area (Dettò et al., 2008). Yet, we argue that the influences should be marginal on the final evaluation results, as our analyses were conducted using monthly footprint climatologies, aggregating footprints from many time steps and different wind directions. The aggregation process tends to cancel out random differences in the lateral dimensions of the half-hourly footprints. To facilitate an in-depth footprint analysis at a finer temporal scale (e.g., daily, hourly), we advocate that V\_SIGMA should be calculated and provided when possible.

Our evaluations were constrained by the availability of both flux tower data and gridded land surface characteristics, which may not be comprehensive in all possible situations. We suggest that future research should further investigate the following aspects. First, the footprint climatology, which aggregates footprints from many time steps and provides a panoramic view of footprints over a certain period, smooths out the variable and extreme footprint conditions as seen by the half-hourly footprints (Text S4). We suggest future research can explore the use of half-hourly footprints directly if there is a need to resolve the fine-scale footprint variability (e.g., CH<sub>4</sub> flux; Helbig et al., 2016; Tuovinen et al., 2019) or if the data-model integration can be carried out at a finer temporal scale (e.g., Xu et al., 2017b). Second, both land-cover type and EVI are important land surface characteristics often used in modeling and upscaling tower-based CO<sub>2</sub> fluxes. However, the relation between the land surface characteristics and fluxes may be nonlinear and ecosystem-specific. We suggest that future research could explore other proxies of fluxes (e.g., machine-learning predictions, model outputs from mechanistic models) such as those used in the evaluation of the network-to-region representativeness (Carvalhais et al., 2010; Papale et al., 2015; Sulkava et al., 2011). Third, our preliminary tests found that ~82 sites had mismatches between the site's IGBP classifications and the dominant land-cover obtained from the land-cover data products (e.g., NLCD, Land Cover Canada). Around 59% of the mismatches could be attributed to differences in classifications, especially in patchy and sparsely-vegetated landscapes (e.g., savanna, shrub-grassland, forest-wetland mosaics), disturbed and restored sites (e.g., logged and replanted forests), and wetlands (see Text S3).

However, we found that around 41% of the mismatches were incorrect land-cover classifications. This discrepancy highlights the need for better assessments and improvements to the available land-cover data products (Jin et al., 2019; Wickham et al., 2017; Wickham et al., 2013). Fourth, the selection of land surface characteristics largely depends on the research questions and data availability (Chen et al., 2012; Kim et al., 2006; Reuss-Schmidt et al., 2019), generally focusing on the vegetation and landscape features that constrain the target fluxes. When available, we suggest considering other fine-resolution (i.e., tens of meters or finer) land surface characteristics in the evaluations. Potential land surface characteristics may include, but are not limited to, land surface temperature (e.g., ECOSTRESS) (Fisher et al., 2020), topographic characteristic, canopy structure, and soils (e.g., LiDAR, digital elevation model) (Chasmer et al., 2011; Giannico et al., 2018; Jucker et al., 2018), vegetation composition (Allred et al., 2020; Davidson et al., 2017), fine-scale disturbance (Xin et al., 2013), vegetation indices (e.g., leaf area index, land surface water index, near-infrared reflectance of vegetation, solar-induced fluorescence) (Griebel et al., 2016; Magney et al., 2019; Pasqualotto et al., 2019; Taddeo et al., 2019), repeated digital camera imagery (Richardson et al., 2018), and other spectral retrievals (Serbin et al., 2015; Singh et al., 2015). Potentially, a holistic approach utilizing multivariate metrics could be adopted similar to the previous works evaluating network representativeness (Chu et al., 2017; Hargrove and Hoffman, 2004; Villarreal et al., 2018). Fifth, the footprint aggregation's temporal interval could be tailored to the retrieval frequency of specific land surface characteristics. For example, with the new terrestrial-monitoring satellites (e.g., ECOSTRESS, Sentinel-2) (Belgiu and Csillik, 2018; Fisher et al., 2020) and other airborne platforms (e.g., unmanned aerial vehicle, airborne visible-infrared imaging spectrometer) becoming available (Chapman et al., 2019; Klosterman et al., 2018; Wang et al., 2019), evaluations could be carried out at a finer temporal scale (e.g., days or weeks). Similarly, proximal optical monitoring at the scale of the flux tower footprint can help understand how best to match remote-sensing and flux data in time and space (Gamon, 2015). Sixth, our evaluation focused on the site-level representativeness and chose 1–6 years of available data at all studied sites. With some AmeriFlux sites being operated for 15–30 years and experiencing changes in vegetation compositions and structures, future evaluations could be carried out separately for individual periods or years (Kim et al., 2018). Similar considerations also apply to ecosystems that experience rapid changes in vegetation compositions and structures (e.g., disturbed, intensively managed sites). Last, many remote-sensing datasets are aggregated in time and/or space (Robinson et al., 2017; Xin et al., 2013) or merged from multiple data sources (Allen et al., 2007; Hilker et al., 2009). This adds extra complexity in matching the spatial and temporal scales between flux and remote-sensing data. While beyond the immediate scope of this study, we recommend that future efforts consider the spatiotemporal scales and relate data aggregation of the flux data and the remotely sensed datasets.

## 5. Conclusions

Eddy-covariance flux data are spatiotemporally dynamic. Large-scale flux datasets, such as FLUXNET2015 and AmeriFlux BASE, need to be used and interpreted with "footprint awareness", especially in research and applications intended to benchmark against models and/or data products with explicit spatial information. This includes, but is not limited to, point-scale simulations and spatially-explicit land surface models, remote-sensing-based models, and machine-learning upscaling studies. The extent and direction of the source areas contributing to the measured fluxes at flux tower sites (i.e., flux footprint) vary largely across sites and through time depending on wind direction, measurement heights, underlying surface characteristics, and turbulent states of the atmosphere. As most flux tower sites are located in heterogeneous landscapes, the spatial variations of land surface characteristics and the temporal dynamics of flux footprints jointly lead to the so-called

representativeness issue, i.e., to what spatial extent do the flux measurements taken at individual tower locations reflect the flux conditions at the corresponding model or data-product grid cells. The commonly adopted model-data integration approach that assumes that the flux footprint represents a fixed target area near the tower universally across sites would certainly introduce biases (e.g., 4%–20% in EVI, 6%–20% in the dominant land-cover percentage). These biases' sign and magnitude were site-specific and varied depending on the target area extent and the land surface characteristics. We suggest that the (half-)hourly footprint information or the required variables for footprint calculations should be generated routinely and made available to the modeling and synthesis user communities when possible. In the meantime, our proposed representativeness indices could be used as a guide to identifying sites suitable for specific applications considering the spatial scales of interest and relevant land surface characteristics. Additionally, the representativeness indices provide general guidance for data use for sites with different representativeness levels, recognizing there are uncertainties in these as well. We advocate that future research should explore a hierarchy of multiple approaches to fully address the representativeness issue.

## Declaration of Competing Interest

None

## Acknowledgments

We thank the AmeriFlux site teams for sharing their data and metadata with the network. Funding for these flux sites is acknowledged in the site data DOI, shown in Table S1. This analysis was supported in part by funding provided to the AmeriFlux Management Project by the U.S. Department of Energy's Office of Science under Contract No. DE-AC02-05CH11231. All footprint climatologies, site-level representativeness indices, and monthly EVI and sensor location biases can be accessed via the Zenodo Data Repository (Datasets S1–S6, <http://doi.org/10.5281/zenodo.4015350>).

## References

- Allen, R.G., et al., 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Applications. *Journal of irrigation and drainage engineering* 133 (4), 395–406.
- Allred, B.W., et al., 2020. Improving Landsat predictions of rangeland fractional cover with multitask learning and uncertainty. <https://doi.org/10.1101/2020.06.10.142489> bioRxiv: 2020.06.10.142489.
- Amiro, B.D., 1998. Footprint climatologies for evapotranspiration in a boreal catchment. *Agric For Meteorol* 90 (3), 195–201. [https://doi.org/10.1016/S0168-1923\(97\)00096-8](https://doi.org/10.1016/S0168-1923(97)00096-8).
- Anderson, M., et al., 2018. Field-Scale Assessment of Land and Water Use Change over the California Delta Using Remote Sensing. *Remote Sensing* 10 (6), 889.
- Arriga, N., et al., 2017. Experimental validation of footprint models for eddy covariance CO<sub>2</sub> flux measurements above grassland by means of natural and artificial tracers. *Agric For Meteorol* 242, 75–84. <https://doi.org/10.1016/j.agrformet.2017.04.006>.
- Baldocchi, D.D., Falge, E., Wilson, K., 2001. A spectral analysis of biosphere-atmosphere trace gas flux densities and meteorological variables across hour to multi-year time scales. *Agric For Meteorol* 107 (1), 1–27.
- Baldocchi, D.D., Krebs, T., Leclerc, M.Y., 2005. Wet/dry Daisyworld': a conceptual tool for quantifying the spatial scaling of heterogeneous landscapes and its impact on the subgrid variability of energy fluxes. *Tellus B: Chemical and Physical Meteorology* 57 (3), 175–188. <https://doi.org/10.3402/tellusb.v57i3.16538>.
- Belgiu, M. and Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204: 509–523. DOI:10.1016/j.rse.2017.10.005.
- Bonan, G.B., Oleson, K.W., Fisher, R.A., Lasslop, G., Reichstein, M., 2012. Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4. *J Geophys Res-Biogeo* 117, G02026.
- Buotte, P.C., et al., 2019. Near-future forest vulnerability to drought and fire varies across the western United States. *Global change Biol* 25 (1), 290–303.
- Carvalhais, N., et al., 2010. Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula. *Biogeosciences* 7 (11), 3707–3729. <https://doi.org/10.5194/bg-7-3707-2010>.

Chaney, N.W., Herman, J.D., Ek, M.B., Wood, E.F., 2016. Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *Journal of Geophysical Research: Atmospheres* 121 (13), 218–235.

Chapman, J.W., et al., 2019. Spectral and radiometric calibration of the next generation airborne visible infrared spectrometer (AVIRIS-NG). *Remote Sensing* 11 (18), 2129.

Chasmer, L., et al., 2011. Characterizing vegetation structural and topographic characteristics sampled by eddy covariance within two mature aspen stands using lidar and a flux footprint model: Scaling to MODIS. *Journal of Geophysical Research: Biogeosciences* 116 (G2), G02026.

Chen, B., et al., 2011. Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling. *Agric For Meteorol* 151 (1), 87–100.

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. DOI:10.1016/j.rse.2012.06.007.

Chen, B., et al., 2010. A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling. *Biogeosciences* 7 (9), 2943.

Chen, L., Dirmeyer, P.A., Guo, Z., Schultz, N.M., 2018. Pairing FLUXNET sites to validate model representations of land-use/land-cover change. *Hydrol Earth Syst Sc* 22 (1), 111.

Chu, H., Baldocchi, D.D., John, R., Wolf, S., Reichstein, M., 2017. Fluxes All of the Time? A Primer on the Temporal Representativeness of FLUXNET. *Journal of Geophysical Research: Biogeosciences*, 122, 289–307. <https://doi.org/10.1002/2016JG003576>.

Chu, H., et al., 2018. Temporal dynamics of aerodynamic canopy height derived from eddy covariance momentum flux data across North American Flux Networks. *Geophys Res Lett* 45, 9275–9287. <https://doi.org/10.1029/2018GL079306>.

Cui, W., Chui, T.F.M., 2019. Temporal and spatial variations of energy balance closure across FLUXNET research sites. *Agric For Meteorol* 271, 12–21. <https://doi.org/10.1016/j.agrformet.2019.02.026>.

Davidson, S.J., et al., 2017. Upscaling CH4 Fluxes Using High-Resolution Imagery in Arctic Tundra Ecosystems. *Remote Sensing* 9 (12), 1227.

Dettò, M., Katul, G.G., Siqueira, M., Juang, J.-Y., Stoy, P., 2008. The structure of turbulence near a tall forest edge: The backward-facing step flow analogy revisited. *Ecol Appl* 18 (6), 1420–1435. <https://doi.org/10.1890/06-0920.1>.

DuBois, S., et al., 2018. Using imaging spectroscopy to detect variation in terrestrial ecosystem productivity across a water-stressed landscape. *Ecol Appl* 28 (5), 1313–1324.

Dumontier, P., Aubinet, M., Lebeau, F., Naikeni, A., Heinesch, B., 2019. Point source emission estimation using eddy covariance: Validation using an artificial source experiment. *Agric For Meteorol* 266–267, 148–156. <https://doi.org/10.1016/j.agrformet.2018.12.012>.

Durden, D.J., Metzger, S., Chu, H., Collier, N., Davis, K.J., Desai, A.R., Kumar, J., Wieder, W.R., Xu, M., Hoffman, F.M., 2020. Automated integration of continental-scale observations in near-real time for simulation and analysis of biosphere-atmosphere interactions. In: Nichols, J., et al. (Eds.), *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*. Springer.

Fisher, J.B., et al., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. *Water Resour Res* 56 (4), e2019WR026058. <https://doi.org/10.1029/2019wr026058>.

Fu, D., et al., 2014. Estimating landscape net ecosystem exchange at high spatial-temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements. *Remote Sensing of Environment* 141, 90–104. <https://doi.org/10.1016/j.rse.2013.10.029>.

Gamon, J.A., 2015. Reviews and Syntheses: optical sampling of the flux tower footprint. *Biogeosciences* 12 (14), 4509–4523. <https://doi.org/10.5194/bg-12-4509-2015>.

Giannico, V., et al., 2018. Contributions of landscape heterogeneity within the footprint of eddy-covariance towers to flux measurements. *Agric For Meteorol* 260–261, 144–153. <https://doi.org/10.1016/j.agrformet.2018.06.004>.

Göckede, M., et al., 2008. Quality control of CarboEurope flux data &ndash; Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems. *Biogeosciences* 5 (2), 433–450. <https://doi.org/10.5194/bg-5-433-2008>.

Griebel, A., et al., 2016. Effects of inhomogeneities within the flux footprint on the interpretation of seasonal, annual, and interannual ecosystem carbon exchange. *Agric For Meteorol* 221, 50–60. <https://doi.org/10.1016/j.agrformet.2016.02.002>.

Griebel, A., Metzen, D., Pendall, E., Burba, G., Metzger, S., 2020. Generating spatially robust carbon budgets from flux tower observations. *Geophys Res Lett* 47 (3), e2019GL085942. <https://doi.org/10.1029/2019gl085942>.

Hargrove, W., Hoffman, F., 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environ Manage* 34, S39–S60. <https://doi.org/10.1007/s00267-003-1084-0>.

Hargrove, W.W., Hoffman, F.M., Law, B.E., 2003. New analysis reveals representativeness of the AmeriFlux network. *Eos* 84 (48), 529–535. <https://doi.org/10.1029/2003EO480001>.

Heidbach, K., Schmid, H.P., Mauder, M., 2017. Experimental evaluation of flux footprint models. *Agric For Meteorol* 246, 142–153. <https://doi.org/10.1016/j.agrformet.2017.06.008>.

Heinsch, F.A., et al., 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *Geoscience and Remote Sensing, IEEE Transactions on* 44 (7), 1908–1925.

Helbig, M., et al., 2016. Regional atmospheric cooling and wetting effect of permafrost thaw-induced boreal forest loss. *Global Change Biol* 22, 4048–4066. <https://doi.org/10.1111/gcb.13348>.

Hilker, T., et al., 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8): 1613–1627. DOI:10.1016/j.rse.2009.03.007.

Hoffman, F., et al., 2017. International land model benchmarking (ILAMB) 2016 Workshop Report. US Department of Energy, Office of Science, 10: 1330803.

Hsieh, C.-I., Katul, G., Chi, T.-w., 2000. An approximate analytical model for footprint estimation of scalar fluxes in thermally stratified atmospheric flows. *Advances in Water Resources* 23 (7), 765–772.

Huete, A., et al., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2): 195–213. DOI:10.1016/s0034-4257(02)00096-2.

Jin, S., et al., 2019. Overall methodology design for the United States national land cover database 2016 products. *Remote Sensing* 11 (24), 2971.

Jucker, T., et al., 2018. Canopy structure and topography jointly constrain the microclimate of human-modified tropical landscapes. *Global change Biol* 24 (11), 5243–5258. <https://doi.org/10.1111/gcb.14415>.

Jung, M., et al., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data* 6 (1), 74. <https://doi.org/10.1038/s41597-019-0076-8>.

Jung, M., Reichstein, M., Bondeau, A., 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.

Kahle, D., Wickham, H., 2013. ggmap: Spatial Visualization with ggplot2. *The R journal*, 5 (1), 144–161.

Katul, G., et al., 2001. Multiscale analysis of vegetation surface fluxes: from seconds to years. *Advances in Water Resources* 24 (9–10), 1119–1132.

Kim, J., et al., 2006. Upscaling fluxes from tower to landscape: Overlaying flux footprints on high-resolution (IKONOS) images of vegetation cover. *Agric For Meteorol* 136 (3–4), 132–146.

Kim, J., Hwang, T., Schaaf, C.L., Kljun, N., Munger, J.W., 2018. Seasonal variation of source contributions to eddy-covariance CO2 measurements in a mixed hardwood-conifer forest. *Agric For Meteorol* 253–254, 71–83. <https://doi.org/10.1016/j.agrformet.2018.02.004>.

Kljun, N., Calanca, P., Rotach, M.W., Schmid, H.P., 2015. A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). *Geosci. Model Dev.* 8 (11), 3695–3713. <https://doi.org/10.5194/gmd-8-3695-2015>.

Kljun, N., Rotach, M.W., Schmid, H.P., 2002. A Three-Dimensional Backward Lagrangian Footprint Model For A Wide Range Of Boundary-Layer Stratifications. *Bound-Lay Meteorol* 103 (2), 205–226. <https://doi.org/10.1023/A:1014556300021>.

Klosterman, S., et al., 2018. Fine-scale perspectives on landscape phenology from unmanned aerial vehicle (UAV) photography. *Agric For Meteorol* 248 (Supplement C), 397–407. <https://doi.org/10.1016/j.agrformet.2017.10.015>.

Kormann, R., Meixner, F., 2001. An analytical footprint model for non-neutral stratification. *Bound-Lay Meteorol* 99 (2), 207–224. <https://doi.org/10.1023/a:1018991015119>.

Kumari, S., Kambhammettu, B.V.N.P., Niyogi, D., 2020. Sensitivity of analytical flux footprint models in diverse source-receptor configurations: A field experimental study. *Journal of Geophysical Research: Biogeosciences* 125, e2020JG005694. <https://doi.org/10.1029/2020JG005694>.

Lamigueiro, O.P. and Hijmans, R., 2018. rasterVis. R package version 0.43.

Latifovic, R., Pouliot, D., Olthof, I., 2017. Circa 2010 Land Cover of Canada: Local Optimization Methodology and Product Development. *Remote Sensing* 9 (11), 1098.

Law, B.E., et al., 2018. Land use strategies to mitigate climate change in carbon dense temperate forests. *Proceedings of the National Academy of Sciences* 115 (14), 3663–3668.

Leclerc, M., Thurtell, G., 1990. Footprint prediction of scalar fluxes using a Markovian analysis. *Bound-Lay Meteorol* 52 (3), 247–258.

Legendre, P., 2014. lmodel2: Model II Regression. R package version 1.7-2. Available at: <http://CRAN.R-project.org/package=lmodel2>.

Loveland, T.R., et al., 1999. An analysis of the IGBP global land-cover characterization process. *Photogrammetric engineering and remote sensing* 65, 1021–1032.

Magney, T.S., et al., 2019. Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *Proceedings of the National Academy of Sciences* 116 (24), 11640–11645. <https://doi.org/10.1073/pnas.1900278116>.

Mauder, M., Foken, T., Cuxart, J., 2020. Surface-Energy-Balance Closure over Land: A Review. *Bound-Lay Meteorol*. <https://doi.org/10.1007/s10546-020-00529-6>.

McMahon, G., Wikem, E.B., Gauthier, D.A., 2004. Toward a Scientifically Rigorous Basis for Developing Mapped Ecological Regions. *Environ Manage* 34 (1), S111–S124. <https://doi.org/10.1007/s00267-004-0170-2>.

Metzger, S., 2018. Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations. *Agric For Meteorol* 255 (28), 68–80. <https://doi.org/10.1016/j.agrformet.2017.08.037>.

Metzger, S., et al., 2013. Spatially explicit regionalization of airborne flux measurements using environmental response functions. *Biogeosciences* 10 (4), 2193–2217.

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

Nicolini, G., et al., 2017. Performance of eddy-covariance measurements in fetch-limited applications. *Theor Appl Climatol* 127 (3), 829–840. <https://doi.org/10.1007/s00704-015-1673-x>.

Omernik, J.M., 2004. Perspectives on the nature and definition of ecological regions. *Environ Manage* 34 (1), S27–S38.

Papale, D., et al., 2015. Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *Journal of Geophysical Research: Biogeosciences* 120 (10), 1941–1957. <https://doi.org/10.1002/2015JG002997>.

Papale, D., et al., 2006. Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences* 3 (4), 571–583.

Pasqualotto, N., Delegido, J., Van Wittenbergh, S., Rinaldi, M., Moreno, J., 2019. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors (Basel)* 19 (4), 904. <https://doi.org/10.3390/s19040904>.

Pastorello, G., et al., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data* 7 (1), 225. <https://doi.org/10.1038/s41597-020-0534-3>.

Pastorello, G.Z., et al., 2017. A new data set to keep a sharper eye on land-air exchanges. *Eos* 98. <https://doi.org/10.1029/2017EO071597>.

R Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria.

Ran, Y., et al., 2016. Spatial representativeness and uncertainty of eddy covariance carbon flux measurements for upscaling net ecosystem productivity to the grid scale. *Agric For Meteorol* 230, 114–127.

Rannik, Ü., Sogachev, A., Foken, T., Göckede, M., Kljun, N., Leclerc, M.Y., Vesala, T., 2012. Chapter 8 Footprint Analysis. In: Aubinet, M., Vesala, T., Papale, D. (Eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer, Dordrecht, pp. 211–262.

Rebmann, C., et al., 2005. Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling. *Theor Appl Climatol* 80 (2–4), 121–141. <https://doi.org/10.1007/s00704-004-0095-y>.

Reuss-Schmidt, K., et al., 2019. Understanding spatial variability of methane fluxes in Arctic wetlands through footprint modelling. *Environ Res Lett* 14 (12), 125010. <https://doi.org/10.1088/1748-9326/ab4d32>.

Ricciuto, D., Sargsyan, K., Thornton, P., 2018. The impact of parametric uncertainties on biogeochemistry in the E3SM land model. *Journal of Advances in Modeling Earth Systems* 10 (2), 297–319.

Richardson, A.D., et al., 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data* 5, 180028. <https://doi.org/10.1038/sdata.2018.28>.

Robinson, N.P., et al., 2017. A Dynamic Landsat Derived Normalized Difference Vegetation Index (NDVI) Product for the Conterminous United States. *Remote Sensing* 9 (8), 863.

Robinson, N.P., et al., 2018. Terrestrial primary production for the conterminous United States derived from Landsat 30 m and MODIS 250 m. *Remote Sensing in Ecology and Conservation*, 4(3): 264–280. DOI:10.1002/rse.274.

Running, S.W., et al., 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment* 70 (1), 108–127.

Schmid, H.P., 1997. Experimental design for flux measurements: matching scales of observations and fluxes. *Agric For Meteorol* 87 (2–3), 179–200.

Schmid, H.P., 2002. Footprint modeling for vegetation atmosphere exchange studies: a review and perspective. *Agric For Meteorol* 113 (1–4), 159–183.

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

Schuepp, P., Leclerc, M., MacPherson, J., Desjardins, R., 1990. Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation. *Bound-Lay Meteorol* 50 (1–4), 355–373.

Serbin, S.P., et al., 2015. Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. *Remote Sensing of Environment*, 167: 78–87. DOI:10.1016/j.rse.2015.05.024.

Singh, A., Serbin, S.P., McNeil, B.E., Kingdon, C.C., Townsend, P.A., 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecol Appl* 25 (8), 2180–2197. <https://doi.org/10.1890/14-2098.1>.

Stoy, P., et al., 2009. Biosphere-atmosphere exchange of CO<sub>2</sub> in relation to climate: a cross-biome analysis across multiple time scales. *Biogeosciences* 6, 2297–2312.

Stoy, P.C., et al., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agric For Meteorol* 171–172 (0), 137–152. <https://doi.org/10.1016/j.agrformet.2012.11.004>.

Sulkava, M., Luyssaert, S., Zaeble, S., Papale, D., 2011. Assessing and improving the representativeness of monitoring networks: The European flux tower network example. *Journal of Geophysical Research: Biogeosciences* 116 (G3), 2156–2202. <https://doi.org/10.1029/2010JG001562>.

Taddeo, S., Dronova, I., Depsky, N., 2019. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote Sensing of Environment* 234, 111467. <https://doi.org/10.1016/j.rse.2019.111467>.

Thomson, D., 1987. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of fluid mechanics* 180, 529–556.

Tramontana, G., et al., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* 2016 (13), 4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>.

Tuovinen, J.-P., Aurela, M., Hatakka, J., Räsänen, A., Virtanen, T., Mikola, J., Ivakhov, V., Kondratyev, V., Laurila, T., 2019. Interpreting eddy covariance data from heterogeneous Siberian tundra: land-cover-specific methane fluxes and spatial representativeness. *Biogeosciences* 16 (2), 255–274. <https://doi.org/10.5194/bg-16-255-2019>.

Verma, M., et al., 2015. Improving the performance of remote sensing models for capturing intra- and inter-annual variations in daily GPP: An analysis using global FLUXNET tower data. *Agric For Meteorol* 214–215, 416–429. <https://doi.org/10.1016/j.agrformet.2015.09.005>.

Villarreal, S., et al., 2018. Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. *Agric For Meteorol* 262, 423–433. <https://doi.org/10.1016/j.agrformet.2018.07.016>.

Villarreal, S., Guevara, M., Alcaraz-Segura, D., Vargas, R., 2019. Optimizing an Environmental Observatory Network Design Using Publicly Available Data. *Journal of Geophysical Research: Biogeosciences* 124 (7), 1812–1826. <https://doi.org/10.1029/2018JG004714>.

Wagle, P., Gowda, P.H., Neel, J.P.S., Northup, B.K., Zhou, Y., 2020. Integrating eddy fluxes and remote sensing products in a rotational grazing native tallgrass prairie pasture. *Science of The Total Environment* 712, 136407. <https://doi.org/10.1016/j.scitotenv.2019.136407>.

Wang, H., Jia, G., Zhang, A., Miao, C., 2016. Assessment of Spatial Representativeness of Eddy Covariance Flux Data from Flux Tower to Regional Grid. *Remote Sensing* 8 (9), 742.

Wang, S., et al., 2019. High spatial resolution monitoring land surface energy, water and CO<sub>2</sub> fluxes from an Unmanned Aerial System. *Remote Sensing of Environment*, 229: 14–31. DOI:10.1016/j.rse.2019.03.040.

Wang, W., Davis, K.J., Cook, B.D., Butler, M.P., Ricciuto, D.M., 2006. Decomposing CO<sub>2</sub>fluxes measured over a mixed ecosystem at a tall tower and extending to a region: A case study. *Journal of Geophysical Research: Biogeosciences* 111 (G2). <https://doi.org/10.1029/2005jg000093>. G02005.

Wickham, J., et al., 2017. Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). *Remote Sensing of Environment* 191, 328–341. <https://doi.org/10.1016/j.rse.2016.12.026>.

Wickham, J.D., et al., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment* 130, 294–304. <https://doi.org/10.1016/j.rse.2012.12.001>.

Williams, I.N., Lee, J.M., Tadić, J., Zhang, Y., Chu, H., 2020. Modeling Spatial Heterogeneity in Surface Turbulent Heat Flux in the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres* 125 (13). <https://doi.org/10.1029/2019jd032255>.

Williams, M., et al., 2009. Improving land surface models with FLUXNET data. *Biogeosciences* 6 (7), 1341–1359.

Xiao, J., et al., 2014. Data-driven diagnostics of terrestrial carbon dynamics over North America. *Agric For Meteorol* 197 (0), 142–157. <https://doi.org/10.1016/j.agrformet.2014.06.013>.

Xiao, J., et al., 2011. Assessing net ecosystem carbon exchange of U.S. terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations. *Agric For Meteorol* 151 (1), 60–69.

Xin, Q., Olofsson, P., Zhu, Z., Tan, B., Woodcock, C.E., 2013. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. *Remote Sensing of Environment* 135, 234–247.

Xu, F., et al., 2017a. Area-averaged evapotranspiration over a heterogeneous land surface: aggregation of multi-point EC flux measurements with a high-resolution land-cover map and footprint analysis. *Hydrol Earth Syst Sc* 21 (8), 4037.

Xu, K., Metzger, S., Desai, A.R., 2017b. Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions. *Agric For Meteorol* 232, 10–22. <https://doi.org/10.1016/j.agrformet.2016.07.019>.

Xu, K., Sühring, M., Metzger, S., Durden, D., Desai, A.R., 2020. Can Data Mining Help Eddy Covariance See the Landscape? A Large-Eddy Simulation Study. *Bound-Lay Meteorol* 176 (1), 85–103. <https://doi.org/10.1007/s10546-020-00513-0>.

Yang, L., et al., 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing* 146, 108–123. <https://doi.org/10.1016/j.isprsjprs.2018.09.006>.

Yang, Y., et al., 2020. Investigating impacts of drought and disturbance on evapotranspiration over a forested landscape in North Carolina, USA using high spatiotemporal resolution remotely sensed data. *Remote Sensing of Environment* 238, 111018.

Yang, Y., et al., 2017. Daily Landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA, using multi-satellite data fusion. *Hydrol Earth Syst Sc* 21 (2), 1017–1037.

Zhang, Y., et al., 2017. A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. *Scientific Data* 4, 170165. <https://doi.org/10.1038/sdata.2017.165>.