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

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

a Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
b Department of Earth System Science, Stanford University, Stanford, CA, USA
c National Ecological Observatory Network Program, Battelle, Boulder, CO, USA
d Department of Atmospheric and Oceanic Sciences, University of Wisconsin Madison, Madison, WI, USA
e Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
f School of Geography and Earth Sciences, McMaster University, Ontario, Canada
g Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, USA
h Department of Environmental Sciences, Policy, and Management, University of California, Berkeley, CA, USA
i USDA-ARS Global Change and Photosynthesis Research Unit
j Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, Canada
k Department of Geography, University of Colorado, Boulder, CO, USA
l Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, OH, USA
m School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA
n School of Environmental Sciences, University of Guelph, Guelph, Ontario, Canada
o Department of Geography & Atmospheric Science, University of Kansas, Lawrence, KS, USA
p Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI, USA
q Earth System Measurements & Data, Pacific Northwest National Laboratory, Richland, WA, USA
r USDA Forest Service, Northern Research Station, New Lisbon, NJ, USA
s Department of Biology, University of New Mexico, Albuquerque, NM, USA
t National Research Council, Institute of Bioeconomy, Rome, Italy
u Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA, USA
v Department of Earth and Atmospheric Sciences and Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada
w School of Natural Resources, University of Nebraska - Lincoln, NE, USA
x Department of Biology, Virginia Commonwealth University, Richmond, VA, USA
y Department of Soil, Water, and Climate, University of Minnesota, Saint. Paul, MN, USA
z Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada
a Northern Research Station, USDA Forest Service, Durham, NH, USA
b Department of Geography and Environmental Studies, Carleton University, Ottawa, Ontario, Canada
c Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization, Tsukuba, Japan
d Department of Environmental Science, Shinshu University, Matsumoto, Japan

* Corresponding author.

https://doi.org/10.1016/j.agrformet.2021.108350

Available online 14 February 2021
0168-1923/Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
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 to identify key model shortcomings and guide their development. 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., Clasmer 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; Gökcede 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., Heirsch 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 $10^2 m^2$ (Hoffman et al., 2017; Williams et al., 2009) or $10^3 m^2$ 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 $10^3$–$10^5 m^2$. 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., $\sim 10^2$–$10^3 m^2$) to represent the conditions at individual sites (Riccuto 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 $10^2$–$10^3 m^2$, 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 (Yin et al., 2013). Typically, a fixed set of pixels near the tower locations are selected (Heirsch 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., $\sim 10^2$–$10^3 m^2$) 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 FLUXNET2015 (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; Gökcede et al., 2008; Rehmann et al., 2005; Wang et al., 2016), none were carried out extensively across a large number of sites, e.g., FLUXNET-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, 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.
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 m s⁻¹) and an independent dataset (R^2 = 0.77, MAE = 0.16 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
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 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 (Kijun 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 (S₀₈₀_piece, S₀₈₀_night (daytime, nighttime)) was defined by the maximal distance from the tower location to the 80% footprint climatology contour. The footprint area (A₀₈₀_piece, A₀₈₀_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 (S₀₈₀_piece, S₀₈₀_night) was calculated using the following equation:

\[
S_{0.80,\text{day}} = \frac{A_{0.80,\text{day}}}{\pi \times (A_{0.80,\text{day}})^2}
\]

Equation 1 applies equivalently to S₀₈₀_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 (O₀₈₀,sea-son_day, O₀₈₀,season_night) and daytime-nighttime overlap index (O₀₈₀,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.

\[
O_{0.80,\text{season-day}} = \frac{1}{K} \sum_{k=1}^{K} \left( \prod_{i=1}^{I} \phi_{ik}^{\text{day}} \right)^{1/k}
\]

\[
O_{0.80,\text{day-night}} = \frac{1}{K} \sum_{k=1}^{K} \left( \sum_{i=1}^{I} \left( \phi_{i1}^{\text{day}} \times \phi_{ik}^{\text{night}} \right) \right)^{1/2}
\]

where i and k designate one pixel and one month, \( \phi_{ik}^{\text{day}} \) and \( \phi_{ik}^{\text{night}} \) 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., \( \sum_{k=1}^{K} \phi_{ik}^{\text{day}} = 1 \)). 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 CO₂ fluxes and as a surrogate of CO₂ 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):

\[
EVI = 2.5 \times \left( \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + 6 \times \rho_{\text{RED}} - 7.5 \times \rho_{\text{BLUE}} + 1} \right)
\]

where \( \rho_{\text{NIR}}, \rho_{\text{RED}}, \) and \( \rho_{\text{BLUE}} \) 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 \( P_{\text{footprint}}: P_{\text{target}} \) 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 ≥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 (\( p > 0.05 \)) between the footprint and a target area (see an example in Figure S8a–S8b).
- **Medium representativeness:** A site had ≥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 (\( p > 0.05 \)) 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 \( P_{\text{footprint}}, P_{\text{target}}, \) 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 \( (EVI_{\text{target}}) \), footprint-weighted EVI \( (EVI_{\text{footprint}}) \), and the sensor location bias \( (\Delta) \) after Schmid and Lloyd (1999):

\[
EVI_{\text{footprint}} = \frac{1}{j} \sum \phi_j \times EVI_j
\]

\[
\Delta = \frac{EVI_{\text{footprint}} - EVI_{\text{target}}}{EVI_{\text{target}}}
\]

where \( j \) designates a pixel, \( J \) denotes the total number of pixels within the footprint climatology, \( \phi_j \) 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:

\[
EVI_{\text{target}} \sim \beta_0 + \beta_1 \times EVI_{\text{footprint}}
\]

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 \( EVI_{\text{footprint}}, EVI_{\text{target}}, \Delta \), 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.klju.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 (model2 package) was adopted for the comparison of EVI and evaluation of V.SIGMA prediction (Legendre, 2014). The ggmmap 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 m² in area (25th and 75th percentiles, Figure 3b). In almost all studied site-years (> 95%), the nighttime footprint climatologies extended farther (~45% on average) and covered a larger area (~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-NRI (\( S_{80\_night} = 0.31 \)), 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 ~32–44% of the studied site-years (i.e., \( O_{80\_season\_day,night} < 0.8 \), 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 ≥ 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., ≥ 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 ~6% and ~20% \( (P_{\text{footprint}} > P_{\text{target}}) \) 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.
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 5e, S20e). The other hand, only half of the wetland sites showed high representativeness even for a 250 m target area (Figures 5e, S20e). The other hand, only half of the wetland sites showed high representativeness even for a 250 m target area (Figures 5e, S20e). 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.
suggesting that the flux footprints mostly covered areas with higher EVI than the surroundings. Only marginal differences (~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 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 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 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.
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.
Agricultural and Forest Meteorology 301-302 (2021) 108350

1. Introduction

2. Methods

3. Results

3.1. Footprint-to-target area representativeness

3.1.1. Forest-dominated landscapes

3.1.2. Cropland-dominated landscapes

3.2. Nighttime footprints

3.3. Spatiotemporal representativeness

3.4. Example Cases

4. Discussion

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

5. Conclusion

Table 1

Pooled results of the comparison of footprint-weighted and target-area EVI for all available site-months (N = 3307). Linear regression model: EVI\textsubscript{target} = β₀ + β₁ \times EVI\textsubscript{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 (β₀) | Slope (β₁) | R² | RMSE | MAE |
|-------------|----------------|------------|----|------|-----|
| Daytime footprint |
| 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 |

| Nighttime footprint |
| 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 both the land-cover composition and EVI (Figures S24–S25). 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 (Δ) 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.
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 $3 \times 3$ 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 (Baldocchi 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; Gockede 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 (Durdan et al., 2020). For example, Griebel et al. (2020) proposed a practical approach to account for

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).
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-Schmid 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; Nicollini 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 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 CH4 and nighttime CO2 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, 1996; 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 (KF01 model)) (Text S2, Figure S3). Yet, as the areas with higher footprint weights were largely overlapped between the FFP and KF01 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; Dumontier et al., 2019; Heidbach et al., 2017; Kumari et al., 2020; Nicollini 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 $V_{\text{SIGMA}}$ could introduce additional uncertainties in the footprint calculations, especially in the lateral dimension of the footprint area (Detto 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 $V_{\text{SIGMA}}$ 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., CH4 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 CO2 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., 2016; Pelle et al., 2015; Sulcova et al., 2013). The 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, scrub-grassland, forest-wetland mosaics), disturbed and restored sites (e.g., logged and replanted forests), and wetlands (see Text S2). 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; Gianmico 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; Pasquolotto et al., 2019; Taddeo et al., 2019), repeated digital camera imagery (Richardson et al., 2015), 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 Galnik, 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 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).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2021.108350.

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 (6), 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.
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)00998-8.
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 CO2 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., Leclere, 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 Galnik, 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., Lansdown, 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-Biogeosci 117, G02026.
Brotto, 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.
