Agricultural and Forest Meteorology 232 (2017) 10–22

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions

Ke Xu [a,*], Stefan Metzger [b,c], Ankur R. Desai [a]

[a] University of Wisconsin-Madison, Department of Atmospheric and Oceanic Sciences, Madison, WI, United States
[b] National Ecological Observatory Network, Fundamental Instrument Unit, Boulder, CO, United States
[c] University of Colorado, Institute for Arctic and Alpine Research, Boulder, CO, United States

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 18 July 2016
Accepted 22 July 2016

Keywords:
Upscale
Eddy-covariance
Regional flux map
Land-atmosphere interaction
Footprint bias
Fine spatio-temporal resolution

ABSTRACT

Eddy-covariance measurements are widely used to develop and test parameterizations of land-atmosphere interactions in earth system models. However, a fundamental challenge for model-data comparisons lies in the scale mismatch between the eddy-covariance observations with small (10^{-1}–10^{1} km^2) and temporally varying flux footprint, and the continuous regional-scale (10^{2}–10^{4} km^2) gridded predictions made in simulations. Here, a new approach was developed to project turbulent flux maps at regional scale and hourly temporal resolution using environmental response functions (ERFs). This is based on an approach employed in airborne flux observations, and relates turbulent flux observations to meteorological forcings and surface properties across the flux footprint. In this study, the fluxes of sensible heat, latent heat and CO_2 integrated over a 20×20 km^2 target domain differed substantially from the tower observations in their expected value (+27%, −9%, and −17%) and spatio-temporal variation (−22%, −21%, and −3%, repsectively) ERF systematic uncertainties are bound within −11%, −1.5% and +16%, respectively, indicating that tower location bias might be even more pronounced for heat and CO_2 fluxes than currently detectable. The ERF-projected fluxes showed general agreement with independent observations at a nearby tower location. Lastly, advantages and limitations of ERF compared to other scaling approaches are discussed, and pathways for improving model-data synthesis utilizing the ERF scaling method are pointed out.

1. Introduction

Earth system models (ESMs) have been developed and are widely used to understand impacts of global climate change (Le Quéré et al., 2013a,b, 2014; Dufresne et al., 2013; Collins et al., 2011). However, uncertainty in the Earth's surface energy budget and terrestrial carbon cycle are found to be a dominant constraint for robust climate projections. For example, the uncertainty in terrestrial CO_2 fluxes can lead to variations of a few hundred ppm in atmospheric CO_2 concentration and several degrees in projected surface temperature (Arora et al., 2013). ESMs benefit from evaluation against direct continuous ecosystem observations of sensible heat, latent heat, water vapor and CO_2 exchange (Schwalm et al., 2010; Xiao et al., 2012; Williams et al., 2009). However, comparing observations to models requires careful consideration of the observation uncertainty and representativeness of observations for the model grid scale. Here, we test a specific approach for improving representativeness and estimating corresponding uncertainty of eddy-covariance (EC) flux tower observations of carbon, water, and heat fluxes.

Eddy-covariance observations of these fluxes have been increasingly used to constrain model uncertainty, because they, in theory, provide reliable spatially distributed and temporally continuous observations of surface-atmosphere exchanges (Bonan et al., 2011; Baldocchi et al., 2001). Parameter sensitivities in photosynthetic rates, respiration allocation, and temperature sensitivity of decomposition in models can, in principle, be constrained by flux tower observations (Dietze et al., 2014), especially when autocorrelation of time series are taken into account (Desai et al., 2010; Desai, 2014). Recently, large model-to-tower syntheses, as part of the North American Carbon Program, have found limitations in modeled spring phenology (Richardson et al., 2012), light use efficiency (Schaefer et al., 2012), and drought sensitivity (Schwalm et al., 2010).

* Corresponding author at: Department of Atmospheric and Oceanic Sciences, UW-Madison, 1225W Dayton St, Madison, WI 53706, United States.
E-mail address: kxu35@wisc.edu (K. Xu).

http://dx.doi.org/10.1016/j.agrformet.2016.07.019
0168-1923/

A fundamental challenge for this kind of model-data comparison is the scale mismatch between a small-scale, spatially non-uniform EC flux footprint and the typically larger-scale gridded, continuous predictions made in simulations. The flux measurement footprint typically represents a small fraction (e.g., <10%) of the model grid cell in most ESMs, and the location of this fraction changes with time. Any transient bias that occurs from changes in sampled characteristics with time can bias model-data comparison. For example, respiration flux chambers demonstrate considerable spatial variability (Jacinthe and Lal, 2006). Recent analysis has shown that consideration of footprint for scaling chamber emissions can significantly improve comparison of EC to chamber-scaled fluxes (Budishchev et al., 2014). For EC fluxes, many sites have differing preferred wind magnitude and directions at daytime versus nighttime, making flux observations at daily scale a mix of diurnal cycle and change of flux footprint. Methods to account for this transient footprint bias could aid in removing a potentially large source of systematic uncertainty for EC flux towers.

Two main scaling approaches, process-based and data-driven approaches, have been utilized for scaling tower-observed surface-atmosphere exchange in space and time, each subject to specific limitations. Purely process-based scaling (Wang et al., 2006; Desai et al., 2008, 2010; Xiao et al., 2011) relies on prescribed mechanistic relationships, oftentimes based on laboratory calibrations and far-reaching assumptions, such as functional steady-state and closure of energy or water balances. Purely data-driven scaling (Xiao et al., 2014; Hutjes et al., 2010) minimizes the number of assumptions employed by inferring relationships among observations directly from the available data, but are limited in model robustness and predictive performance, in particular for discrete predictions with substantial intra-class variability (Prueger et al., 2012; Wang et al., 2006). Among many data-driven methods, machine-learning techniques, such as artificial neutral network (Sulkava et al., 2011; Papale and Valentini, 2003) and model tree ensemble (Jung et al., 2010), have been used to regress atmospheric fluxes against surface properties. However, due to a lack of sample size, the temporal resolution of these approaches is typically aggregated to daily to monthly timescales, and cannot provide information on the diurnal cycle. Further consideration of transient footprint bias is usually neglected in either case.

The environmental response function (ERF) approach is a scaling algorithm that combines data-driven and process-based approaches to the sub-hourly timescale, and provides temporally and spatially resolved flux grids (Metzger et al., 2013). The principle underlying ERF is to extract the relationship between observed biophysical drivers and ecological responses using machine learning, with explicit consideration of sub-hourly flux footprint variation. Provided sufficiently good calibration, the resulting ERF in conjunction with spatio-temporally explicit grids of biophysical surface properties can be used for upscaling the surface-atmosphere exchange into larger target areas. Thus far, the ERF approach has been developed with and utilized for aircraft-based EC measurements in the spatial domain (Metzger et al., 2013).

The present study develops and tests an ERF approach for tower-based flux observations, based on the measurements from the AmeriFlux Park Falls WLEF very tall tower in North Wisconsin, USA during July and August 2011. We ask:

(i) Is it possible to map, based on turbulent fluxes from a very tall tower, spatio-temporally explicit flux fields covering a 20 × 20 km^2 regional domain with acceptable uncertainty (<20%, corresponding to a frequently observed energy imbalance (Foken 2008; Stoy et al., 2013))?
(ii) How do ERF-scaled fluxes compare to alternative approaches, e.g. process-based and data-driven upscaling methods, and what are the prospects of using ERF-scaled fluxes for improving ESMs?
(iii) What are the limitations and where is potential for improvement of the current ERF algorithm?

In this paper, we first introduce the climate, biophysical properties in the study area and footprint composition of WLEF (Section 2.1). The methodology of ERF scaling approach and associated uncertainty algorithm are described in Section 2.2 and Section 2.3. We present the extracted ERFs, scaled flux grids and the associated uncertainty budget in Sections 3.1, 3.2 and 3.3, separately. A comparison of ERF scaling approach and its outcomes with in-situ measurements and other scaling approaches is performed in Section 3.4. In Section 3.5, we give a prospect for model-data comparison improvement with ERF scaling approach. Lastly, the current limitations of ERF scaling approach and potential improvements are discussed in Section 3.6, and conclusions in Section 4.

2. Materials and methods

2.1. Study area and data acquisition

The 447-m tall WLEF television tower (45.9°N, 90.3°W) is located in the Mississippi River Basin in the Park Falls Ranger District of the Chequamegon-Nicolet National Forest, Wisconsin, USA (Fig. 1). Small elevation changes, on the order of 20 m, create a mixed landscape of wetlands and upland forests. Previous studies (Desai et al., 2008, 2015) have indicated that the footprint climatology samples a landscape that is representative of much of the Upper Midwest U.S. forested region, and the proportions of wetland and forest sampled are representative of the average wetland/forest coverage in the entire National Forest. The surrounding forest canopy has approximately 70% deciduous and 30% coniferous trees, and a mean canopy height of 20 m. The whole region was heavily logged around the beginning of the 20th century, similar to many forests in the north central United States. Soils are sandy loam and are mostly glacial outwash deposits. The site has an interior continental climate with cold winters and warm summers.

Observations used in this study include tower-measured meteorological state variables and turbulent exchanges, as well as surface properties from satellite remote sensing. Tower-based 10 Hz observations were chosen from 27th July to 20th August 2011 for WLEF at 30 m and 122 m levels. For turbulent sensible heat flux (H), this study utilized 30 m and 122 m EC measurements, but for latent heat (LE) and CO_2 flux only the 122 m data were used due to poor quality of measurements at the lower level. The flux footprint fetch was 0.1–0.8 km and 2–5 km for 30 and 122 m height measurement at 90% cumulative level (Fig. 2a), respectively. During the study period, the dominant land cover types within the flux footprint of 122 m measurement level were woody wetlands (44%), deciduous forest (25%), and mixed forest (9%). Other land cover types (e.g. grassland, cropland) contributed less than 5%.

Dry mole fraction of CO_2 and water vapor were measured by a closed-path infrared gas analyzer (LI-COR, Inc. Lincoln, USA, LI-6262). Fast response wind speed and air temperature were derived from a collocated sonic anemometer (Applied Technogies., Inc. Seattle, USA, ATI Type K). Reference air temperature and relative humidity were also measured (Vaisala, Inc. Louisville, USA, HMP45C). Additional measurements at the surface included the barometric air pressure (Vaisala, Inc. Louisville, USA, PT101B) and the incoming photosynthetic active radiation (PAR) (LI-COR, Inc. Lincoln, USA, LI-190 Quantum Sensor).

Land surface temperature (LST) and enhanced vegetation index (EVI) were chosen as the biophysical surface drivers to approximate the spatial variation in the target domain following Metzger

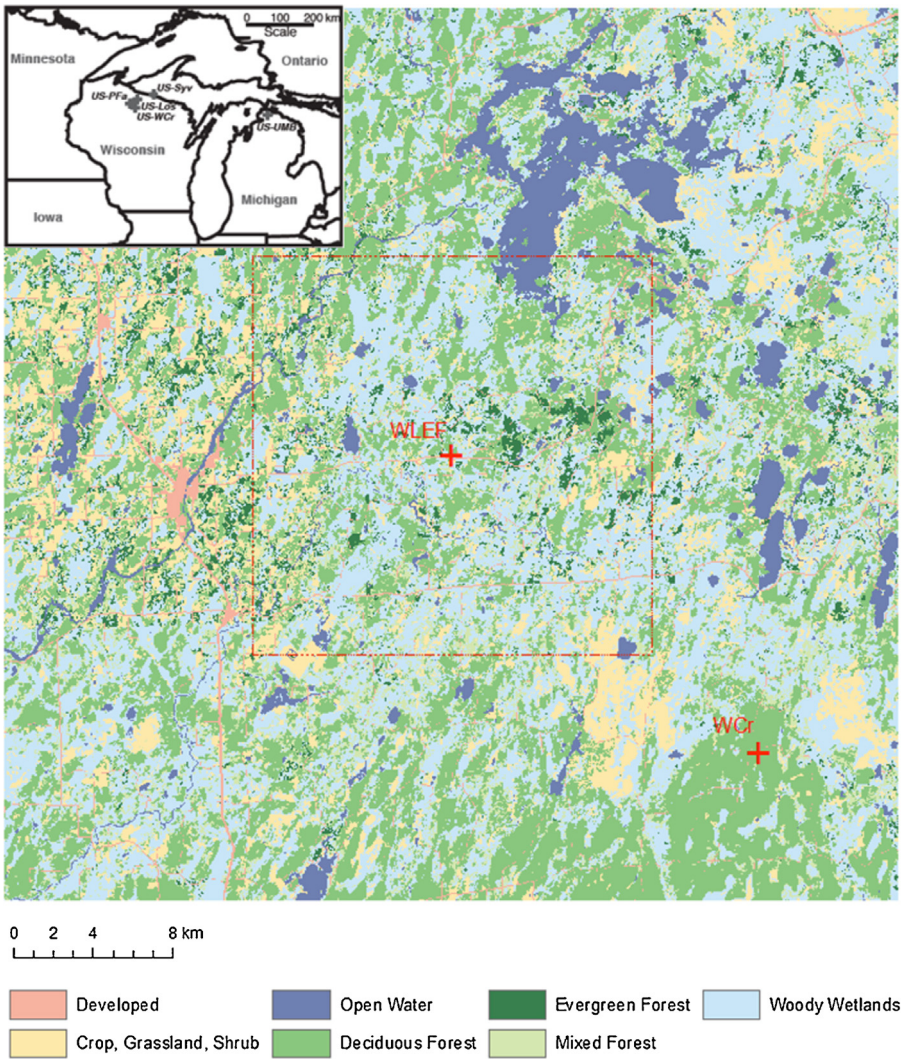

Fig. 1. Land cover map (National Land Cover Database 2011) of Park Fall tower and Willow Creek tower sites (+) and $20 \times 20\,km^2$ target domain (within red dashed line). The land cover data were obtained from National Land Cover Database 2011 (NLCD 2011) at spatial resolution of 30 m. NLCD 2011 is based primarily on a decision-tree classification of Landsat satellite data (Homer et al., 2015). Map of Upper Midwest (USA) is in the up left panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

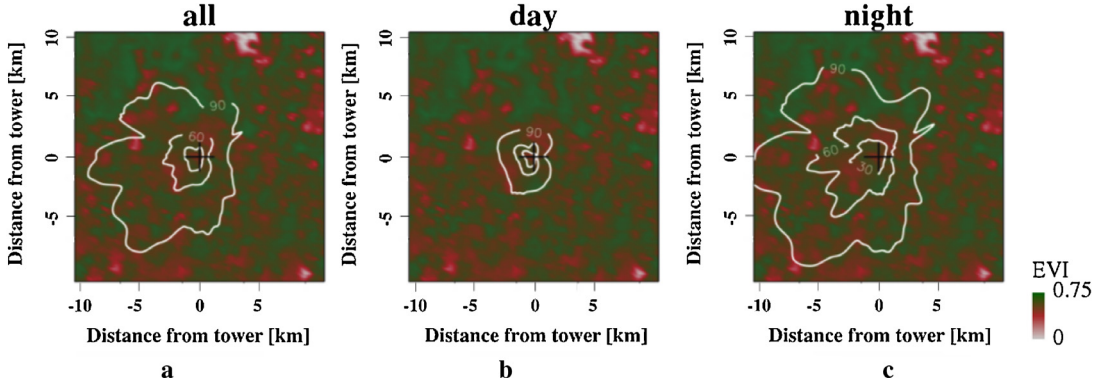

Fig. 2. Footprint climatology (30%, 60% and 90%, white contour lines) for 122 m level measurements superimposed over MODIS enhanced vegetation index (EVI) within $20 \times 20\,km^2$ target domain surrounding the tower, which is indicated with the central crosshairs. (a) entire study time period; (b) daytime (9:00–17:00 CST), and (c) nighttime (17:00–9:00 CST).

et al. (2013). These spatially explicit environmental drivers were acquired from Moderate Resolution Imaging Spectroradiometer (MODIS) data products. 250 m 16-day interval MOD13Q1 (V005) EVI and 1000 m 8-day daytime MYD11A2 (V005) daytime LST were used. Atmospheric boundary layer (ABL) height, z_i, was obtained by linear interpolation into one-minute interval from the North American Regional Reanalysis (NARR) 3-hourly data produced by National Oceanic and Atmospheric Administration (NOAA).

An independent validation was established using EC observations over the same study period from a nearby Ameriflux tower site, Willow Creek (WCr) at the 30 m level. Willow Creek is located 21 km southeast to WLEF-TV site in the upland mature deciduous forests of the Chequamegon-Nicolet National Forest (Fig. 1). The surrounding stand is relatively homogeneous within 0.6 km of the tower. Common geomorphological features of the upland areas include southwest-trending drumlins, slightly elevated ground moraines and poorly drained depressions. These secondary stands are conversions from the old-growth hemlock-hardwood forests to sugar maple-aspen-yellow birch forests. Data and associated uncertainty were calculated using the methodology in Cook et al., (2004). The uncertainty of WCr observations was calculated as the standard deviation of the measured fluxes measured for a sliding window of +/− 5 days and +/− 1 h of the current timestamp, when the meteorological conditions must also be sufficiently similar. In order to detect the systematic difference between WCr-measured and ERF-projected fluxes, Deming regression was used (Ripley and Thompson, 1987), to account for random uncertainty of both observations and projections by assigning a weight to each data inversely proportional to its error variance.

2.2. Environmental response functions (ERF) approach

We adapted the ERF scaling procedure for tower EC measurements based on the airborne ERF scaling approach described in Metzger et al. (2013). The underlying principle of ERF is to utilize high-frequency (minute to minute) footprint variation to extract the relationship between high-frequency flux response and appropriate spatial or temporal drivers, e.g. meteorological forcings and surface ecological properties, and then utilize the extracted relationship for spatio-temporal mapping to the whole domain, where possible. Consider the metaphor of the Blind Man and the Elephant: the tower ("the blind man") can describe only one part of "the elephant" of ecosystem fluxes at a time, namely those fluxes that arise from the flux footprint influence area. If the "elephant" were stationary, then over time, the time-varying footprint would allow the tower to map fluxes without any scaling approach needed. However, in our case the "elephant" moves and changes its behavior, e.g. with the diurnal cycle. Thus, ERF is a data-assimilation approach that attempts to recover the whole picture of "the moving elephant" using snapshots of varying parts of "the elephant" over space and time.

Key processing steps in ERF included: i) quantifying wavelet-decomposed turbulent flux response and footprint-weighted environmental drivers of each flux observation (Sections 2.2.2, 2.2.3, Fig. 4); ii) extracting ERFs between flux responses and drivers using machine learning (Section 2.2.4, Figs. 5 and 6); iii) projecting turbulent exchange grids over the target area based on extracted ERFs and key drivers in each grid cell (Section 2.2.4, Fig. 7). Our routines were developed in GNU R version 3.1 (R Development Core Team, 2012). Algorithm code and examples are being developed for a public repository.

Initially, we determined appropriate time and space-scales for aggregation (Section 2.2.1). Next, high-rate (one-minute) turbulent exchange was calculated using wavelet discretization for tower-based measurements permitting inclusion of transporting scale up to 40 min (Section 2.2.2). The corresponding environmental drivers included two types: temporally explicit meteorological variables and spatially resolved land surface properties (Section 2.2.3). Lastly, ERFs was extracted and used for projection in Section 2.2.4.

2.2.1. Scale considerations

We determined a suitable averaging window for meteorological variables and fluxes as a trade-off between random error and temporal resolution. On the one hand, the random error of a flux is inversely proportional to the square root of the averaging time period (e.g. Lenschow and Stankov, 1986). On the other hand, a high temporal resolution ensures that the spatio-temporal variability of drivers and responses is captured as complete as possible in the ERF training dataset (Fig. 3). For this purpose we estimated vertical and horizontal transit times: Based on convective velocity, the transit times for a surface emission to reach a measurement level are 8.5 ± 7.5 s and 83 ± 74 s for the 30-m and 122-m measurement levels, respectively. Next, based on column-average horizontal velocity, it took 187 ± 40 s for an emission from a patch the size of one characteristic surface length scale (411 ± 88m, e.g. Strunin and Hiyama, 2004) to flow through the tower. With a much longer averaging window, the tendency increases for an observed flux to be a blended signal from different surface patches. On this basis, an averaging window of five minutes was found to be a feasible compromise between random error (Table 1) and spatio-temporal resolution.

2.2.2. Wavelet discretized turbulent exchange

Before wavelet discretization and meteorological state variable calculation, several preprocessing steps were performed building on Metzger et al. (2013): (i) Thresholding for physically unfeasible values and de-spiking after Brock (1986) and Starkenburg et al. (2015); (ii) Planar fit rotation (Wilczak et al., 2001) of the 10 Hz wind data; (iii) Time delay correction using maximum correlation; (iv) Point-by-point conversion of sonic temperature to air temperature based on Schotanus et al. (1983).

With consideration of transporting scales up to 40 min, EC data were processed to obtain fluxes at a one-minute interval following the approach of Metzger et al. (2013) which used a continuous wavelet transform (Torrence and Compo, 1998). In the turbulent flux calculation, flux contributions beyond the cone of influence was considered, because only by this, the difference between standard EC method and wavelet cross-scalogram was within a few percent and avoided biasing the results. Thereafter, a correction for high-frequency spectral loss following Ammann et al. (2006) was directly applied to the wavelet cross-scalograms, respectively. In praxis, according to Section 2.2.1, time series of flux observations were integrated over the cross-scalograms over a five-minute window and the window was moved one-minute forward in time for each observation. The instrument detection limit for fluxes is estimated after Billesbach (2011) in wavelet cross-scalogram, and random and systematic statistical errors of turbulent sampling were calculated after Lenschow and Stankov (1986) and Lenschow et al. (1994).

A list of criteria were implemented to select reliable flux observations for subsequent steps: i) Stationarity and integral turbulence characteristic tests helped avoid observations that do not fulfill fundamental assumptions on EC (Foken and Nappo, 2008, Table 4.2, usability for long-term measurements when test result within 250% of model); ii) To omit potential decoupling and periods of large below-sensor storage fluxes, $u^* > 0.2 ms^{-1}$ was used to select H at 30 m and 122 m levels, $H > -20\,\mathrm{W\,m^{-2}}$ at 30 m was used to select fluxes at 122 m, and $w^* > 0.8 ms^{-1}$ was used to select LE and CO_2 flux to avoid bias from gas exchange with large storage fluxes, since w^* has a negative relationship with storage flux in this study. In result, qualifying sensible heat flux observations spanned the entire diurnal cycle, mainly because data from both, 30-m and 122-m level were available. For LE and CO_2, only daytime data remained.

2.2.3. Land surface and meteorological state variables

Meteorological variables measured by towers provided high temporal resolution, but are of limited spatial representativeness. On the other hand, remote sensing data products could provide spatially resolved land surface properties, but are of limited temporal resolution. Here, we combined these two complementing types of

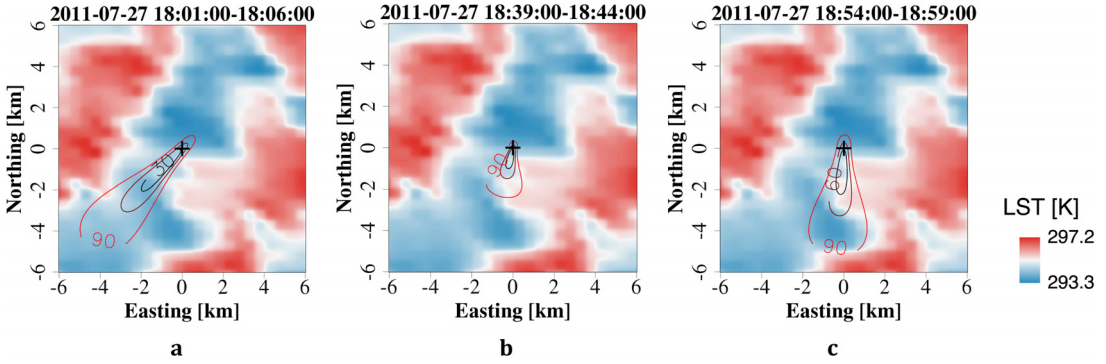

Fig. 3. Flux footprint (30%, 60% and 90%, contour lines) for 122 m level measurements over MODIS-land surface temperature (LST) on July 27th, 2011, (a) 18:01–18:06, (b) 18:39–18:44, (c) 18:54–18:59 CST.

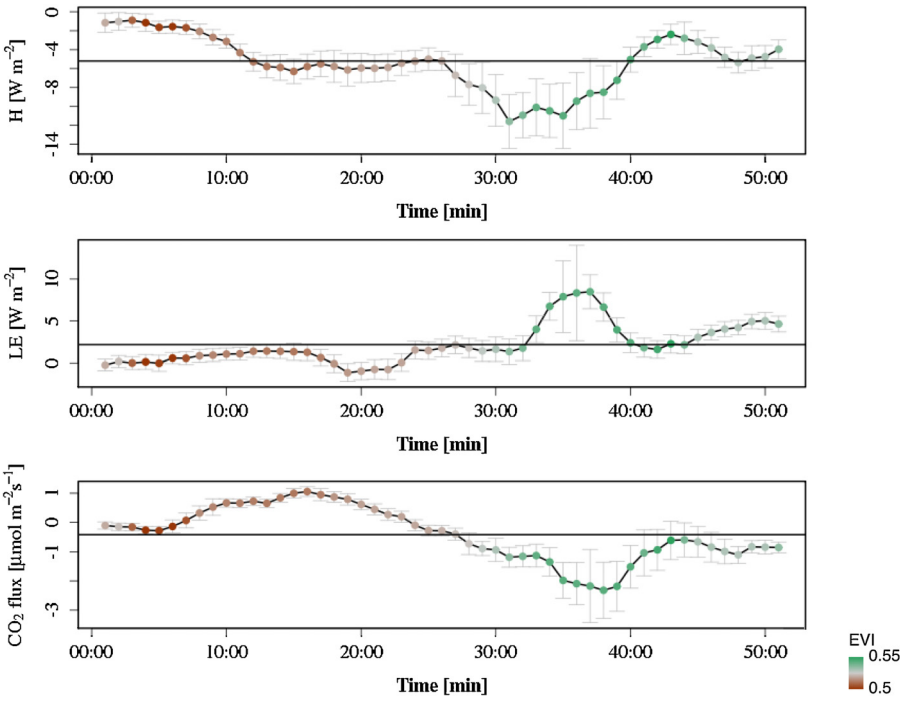

Fig. 4. Time series of turbulent sensible heat (H), latent heat (LE), and CO_2 fluxes on July 27th, 2011, 18:00–19:00 CST. Colors and grey bars indicate the footprint-weighted EVI and 1 sigma random uncertainty of each flux observation, respectively. Black horizontal lines show the hourly-average fluxes calculated from traditional eddy-covariance method. Positive CO_2 fluxes imply carbon exchange from the surface to the atmosphere (and vice versa) in this paper.

Table 1
Median systematic and random uncertainty terms (in parenthesis) of a single flux observation (for (i) and (ii)) or of a single projected cell (for (iii), (iv), (v) and (vi)).

Source	H	LE	CO_2 flux
(i) Instrumentation and hardware	0.9 (1.2) W m^{-2}	1.4 (1.8) W m^{-2}	0.2 (0.3) umol m^{-2} s^{-1}
(ii) Turbulent sampling	1% (45%)	1% (78%)	0.4% (52%)
(iii) ERF state variables	1% (49%)	0.2% (57%)	0.1% (52%)
(iv) Spatio-temporal analysis	0.2% (46%)	−0.4% (58%)	−4% (71%)
(v) ERF training	0.06% (4%)	0.2% (2%)	0.1% (2%)
(vi) ERF projection	**−11% (130%)**	**−1.5% (72%)**	**16% (51%)**

drivers, and used them to explain the temporal and spatial variation of turbulent exchange response.

After preprocessing steps mentioned in Section 2.2.2, meteorological variables were calculated by averaging a centered time-frequency window of five minutes in time. For each subsequent observation, the window moves one minute forward in time (i.e. four minutes overlap with the preceding observation).

All remote sensing data were mosaicked, reprojected and cut into the 20 km × 20 km predefined target area centered around the tower. EVI and LST at 100 m target resolution were obtained by bilinear interpolation. EVI and LST were temporally downscaled to one-hour interval by linear interpolation. Surface properties are more responsible for spatial variability of the response surface in the machine learning algorithm, while meteorological drivers, e.g. solar azimuth, are more responsible for the diurnal cycle and temporal variability of the response. Also considering the long temporal repeat interval (8/16 day) of remotely-sensed surface properties, we chose to use linear interpolation so as to avoid adding artifi-

cial diurnal cycle of surface properties. The uncertainty generated from the unaccounted spatial and temporal variation within the measured spatial and temporal resolution (e.g. 1 km and 8 day) was addressed in the uncertainty budget (Section 2.3, Table 1(iii)).

For each flux observation, the corresponding land surface drivers were quantified by convolving the spatially-explicit driver (e.g. LST) with a footprint weight matrix. For this purpose we used the footprint parameterization of Kljun et al. (2004) in combination with a crosswind distribution (Metzger et al., 2012). The inputs of the resulting parameterization included standard deviation (sd) of the cross-wind σ_v, sd of the vertical wind σ_w, u_*, roughness length z_0, measurement height z_m, of which, σ_v σ_w, u_* were directly

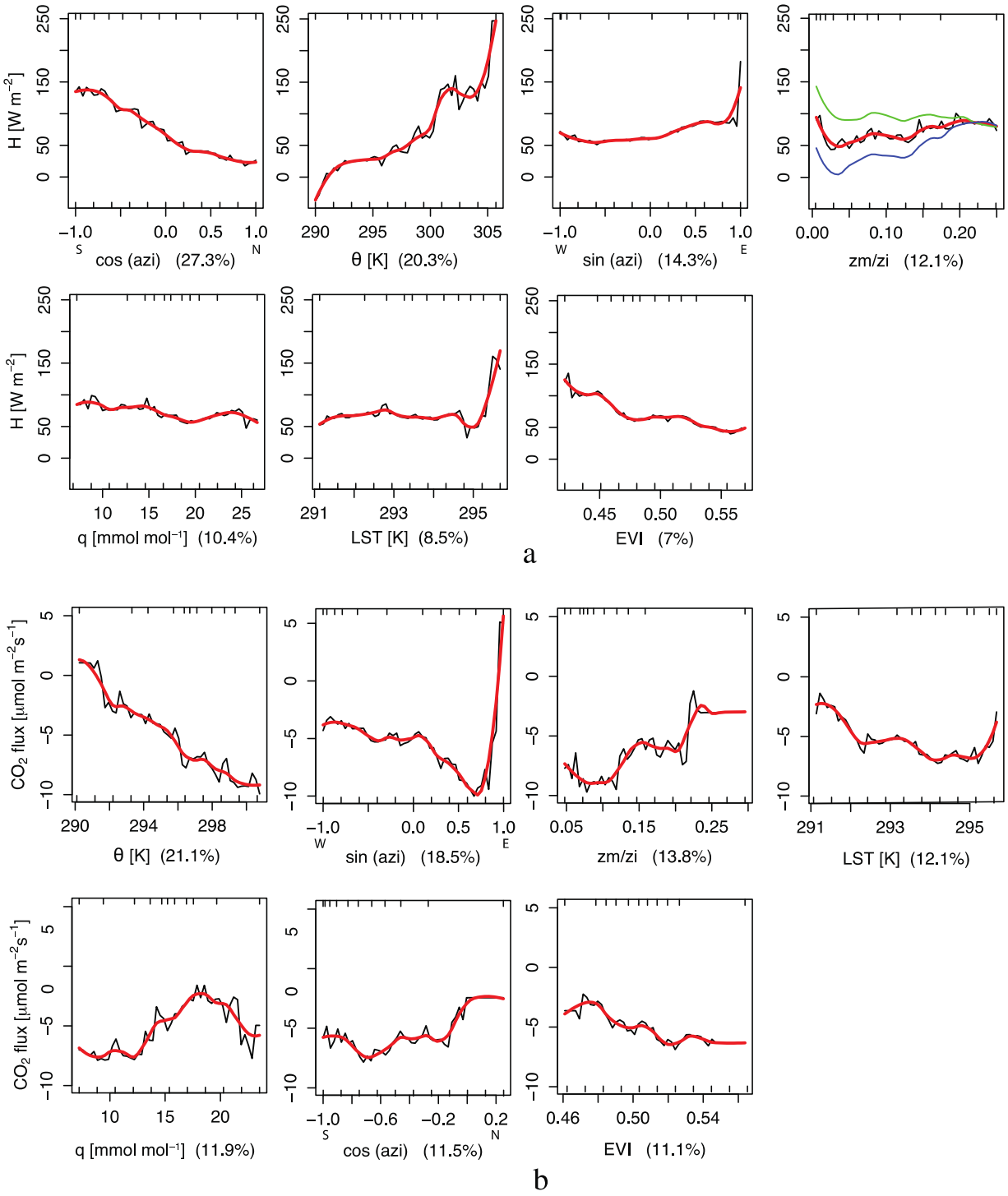

Fig. 5. Equidistant response plots are shown for sensible heat (H, panel a) and CO_2 flux (panel b). The drivers are presented in decreasing order of explained variation in the flux signal (partial R^2 in braces). The black lines are the fitted integrated response over the range of one individual driver. Smoothed representations of the fitted function (locally weighted polynomial regression) are shown in red bold lines. The equidistant response plots are generated using uniformly distributed percentiles within the range of training data shown as inward tickmarks on the lower x-axis. Training data percentiles are shown as inward tickmarks on the upper abscissa. (W), (E) under sin(azi) subplot and (N), (S) under cos(azi) subplot indicate western, eastern, northern and southern direction of the solar azimuth, respectively. The green and blue curves in the subplot of z_m/z_i in (a) show the partial response of H against z_m/z_i integrated separately over daytime (9:00–17:00 CST) and nighttime (17:00–09:00 CST), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

obtained from the tower. z_0 was calculated after Businger et al. (1971) in the form of Högström (1988) with displacement height set as 13 m, which is 2/3 of the mean tree height around the site.

2.2.4. Environmental response function extraction and projection

ERFs were generated by machine learning with observed turbulent exchange (response) and footprint-weighted biophysical surface properties and meteorological states (drivers). Building on Metzger et al. (2013), we used boosted regression trees as machine learning technique, which is based on categorization and regression. It produces a composite model through combining many local models that are regressed at dividing points of the categorization (Zhou, 2012). The extracted ERFs were then summarized in equidistant response-sensitivity plots (Cacuci, 2003), which show the driver-response relationships stored in the ERF (Fig. 5). In each plot, the ERF was evaluated with random combinations of drivers drawn from uniform distributions, and the response was aggregated for each driver individually. The contribution of each driver to explained variance was determined from the reduction of R^2 when removing one driver at a time.

In terms of driver selection, solar azimuth angle, relative measurement height within boundary layer, temperature and water vapor gradients between surface and atmosphere were important environmental effects controlling the observed fluxes. To explain the diurnal cycle and solar radiation, we included cos(azi) and sin(azi), where azi denotes the solar azimuth angle. cos(azi) describes how high the sun is when it is up. Besides, during nighttime, cos(azi) changes its magnitude to indicate midnight or near sunset/sunrise, while solar zenith angle always remains 0. sin(azi) can distinguish mornings (<180°) and afternoons (>180°), when solar zenith angle cannot. The combination of zenith angle and azimuth angle would be useful to express seasonal dynamics over longer periods. Here, the study period is too short to consider such seasonal dynamics, and the zenith angle is not included as a driver. Therefore, for our purposes, solar azimuth angle, θ, can better explain diurnal variability than solar zenith angle, and hence help de-convolve the flux temporal variation from flux spatial variation better. Considering the relative measurement height in the ABL, z_m/z_i, not only combines EC measurements from different heights z_m, but also to explicitly account for vertical flux divergence during ERF projection. Lastly, the vertical gradients of temperature and water vapor permit the ERF to explain corresponding surface-atmosphere exchanges as flux-gradient relationships. Here, air potential temperature (θ in K) and mole fraction of water vapor in dry air (q in $mmolmol^{-1}$) were used as meteorological drivers, as well as LST and EVI as corresponding land surface drivers. While EVI represents the green vegetation fraction, in the absence of a suitable surface moisture product it was chosen as proxy for surface-available moisture in a transpiration-dominated environment.

The resulting ERF acted as a transfer function and was applied to project turbulent exchange to each 100 m grid cell across the $20 \times 20\ km^2$ target domain at minutely interval. The projection was at 30 m height for H and 122 m for LE and CO_2 flux so as to be compared with 30 m H observations at both WLEF, WCr tower sites and 122 m WLEF LE and CO_2 observations. During projection, we used the median value during the time interval for the meteorological drivers, assuming that the atmospheric state above the target area was spatially homogeneous. The uncertainty resulting from this assumption was quantified in the uncertainty budget (Section 2.3, Table 1(iii)). The spatio-temporally resolved grids were used for LST and EVI.

When summarized over the study period, because of the uneven distribution of qualified observations and projections (more observations and projections during daytime than nighttime), the monthly averaged observations and projections were calculated as mean of monthly-mean diurnal cycle, in the cases of latent heat and CO_2 flux excluding nighttime data.

2.3. Uncertainty budget

Throughout the study we use median and median absolute deviation (MAD) for quantifying systematic and random uncertainty, respectively (Croux and Rousseeuw, 1992; Rousseeuw and Verboven, 2002).

To evaluate the significance of the presented approach, we estimated the uncertainty budget throughout the procedure, including the uncertainty from: (i) instrument detection limit, (ii) systematic and random uncertainty for turbulent sampling, (iii) ERF state variable uncertainty, (iv) spatio-temporal analysis, (v) ERF training uncertainty, and (vi) ERF projection uncertainty. Of those, (i) and (ii) have been introduced in Section 2.2.2. To calculate (iii)–(v) we used the routine of Metzger et al. (2013), in short: (iii) quantified the uncertainty generating from unaccounted spatial and temporal variation of state variables, as we assumed tower-observed θ and q were spatially homogeneous, and the spatially resolved drivers were temporally linear continuous in flux map projection. (iv) quantified the uncertainty in the ERF scaling approach resulting from footprint modeling and implicit assumption of linear mixing in machine learning. (v) used random cross-validation (CV) to assess how well ERF projection performed when projecting within the tower footprint originally used in model training. (vi) consisted of an updated approach using stratified CV, in order to evaluate how well ERF performs when projecting to areas the tower footprint had never covered during the training period. This term was essentially important, as the ultimate purpose of the ERF approach was to project to a consistent target area that the tower footprint cannot sample at all times. We divided the target area into a northeastern, southeastern, southwestern and northwestern quadrant. On this basis, four incomplete training datasets were created, each of which omitting all data from one quadrant by wind direction. For each incomplete training dataset, (i) the ERF was trained with data from three quadrants; (ii) the resulting ERFs along with the state variables from the omitted quadrant were used for projection; (iii) The resulting projection was compared to the reference projection not subject to omissions.

3. Results and discussion

In Section 3.1, we first discuss the ERF outcome, the extracted response functions for H and CO_2 flux, based on the high-frequency flux computation of the ERF approach. In Section 3.2, the projection results from ERFs and summarized probability density functions (PDFs) are shown and discussed. Sections 3.3–3.6 present and discuss the uncertainty, comparison with other upscaling approaches, as well as prospect and limitation of ERF scaling, respectively.

3.1. Extracted environmental response functions

This section presents the extracted case study environmental response functions. Flux footprints from tower EC measurements can vary rapidly, e.g. with wind direction and wind speed. In the case of Fig. 3, the footprint changed from southwest narrow strip (Fig. 3a) over small closed area (Fig. 3b) to south larger area (Fig. 3c) within the hour. In consequence, the EC measurement sourced surface areas with potentially very different emission characteristics in rapid succession, or even a blend thereof. Only EC fluxes at high temporal resolution (i.e., one-minute in this study) distinguishes such source area variations through de-convolving what would otherwise be an hourly blend. This provides the foundation for separating spatial (surface characteristics) and temporal

(ecosystem activity) attributes when relating a varying flux signal to environmental drivers.

These wavelet computed fluxes were comparable to traditional hourly non-wavelet based approaches. We found the bias of our computed fluxes, compared to the fluxes calculated from traditional EC method over the whole dataset, were within $-3.6\% \pm 4\%$, $-8.7\% \pm 4\%$ and $+7.5\% \pm 3\%$ for H, LE and CO_2 flux respectively, where tolerance represents one standard error. Visual inspection of wavelet computed fluxes against footprint weight drivers showed the value of high-resolution flux processing. For example, in Fig. 4, between minutes 30 and 40, an oasis effect was observed consistent with land surface wetting and warming in the surface flux footprint. The sensible heat flux became negative as heat was dominantly released by a wet surface through latent heat. This was also indicated by mirrored amount of latent heat flux increase. The advantage to resolve and attribute effects of flux footprint changes on the minute-scale stands out when compared to fluxes calculated from traditional hourly EC method (black lines in Fig. 4). The latter remains constant for the entire duration, and is unable to reflect spatio-temporal variations at finer resolution. The advantage of utilizing wavelet cross-scalogram fluxes was also explicitly confirmed in the earlier publication (Metzger et al., 2013).

Fig. 5 shows non-linear, cross-correlated, multi-dimensional ERFs extracted from BRT. It is noted that the absolute values shown on the y-axis of Fig. 5 do not imply the actual projected flux, as the responses were not projected with the actual driver combination, but random combinations of uniformly distributed samples within the range of the drivers (Section 2.2.4). However, these equidistant plots are most useful for revealing the relationships among driver and response stored in the ERF.

Strong relationships between H and cos(azi), θ, sin(azi), z_m/z_i, LST and EVI have been found in this study as shown in Fig. 5 a, while we did not find significant relationship between H and q. As shown in cos(azi) subplot, large (small) amount of sensible heat was released from surface during daytime (nighttime) with negative (positive) cos(azi). The variability of cos(azi) could roughly capture 1/3 of H variation. It is interesting to note a strong positive relationship between H and θ. This is consistent with our general understanding that the air is warmed by the surface mainly through sensible heat, latent heat and long wave radiation. Another possible explanation could be thatθ acts as a proxy for the actual radiative forcing due to the presence of clouds, but further study is required to test alternative hypotheses. In sin(azi) subplot, the pulse at sin(azi) around 7–9 a.m. is the sign of buoyancy overcoming nighttime stable stratification at the beginning of the convective boundary layer (CBL) development. In the subplot of z_m/z_i in Fig. 6a, the green and blue curves indicate that during daytime, H declined with z_m/z_i, and in nighttime, H was positively related with z_m/z_i. This pattern is also reflected in the cross-relationship between z_m/z_i, cos(azi) and H (Fig. 6a). Next, the ERF of H with q had a small negative slope, following general expectation of higher H during drier periods. Lastly, H was generally positively related to LST and negatively related to EVI, indicating that warmer and less vegetated areas tended to develop larger heat flux. The cross-correlation between EVI, cos(azi) and H displays that during daytime H is larger over areas with lower EVI, whereas it is smaller at lower EVI during nighttime (Fig. 6c). This indicated that drier areas with lower EVI were heating and cooling rapidly, and wetter areas with higher EVI could store more heat during daytime and release it during nighttime.

For CO_2 turbulent flux, θ, sin(azi) and z_m/z_i were the most important drivers. As expected, turbulent CO_2 flux was negatively and positively related with θ and z_m/z_i, respectively. In the sin(azi) subplot, the flush of nocturnal accumulated storage flux in the early morning is captured. The magnitude of CO_2 flux was larger during noontime than morning and afternoon as indicated in the cos(azi) subplot. Areas with higher LST and EVI took up more CO_2, consistent with general understanding. The ERF between q and CO_2 flux was quite non-linear. As expected, the plants could uptake more CO_2 during wetter times. Plants were also shown to uptake more CO_2 at dry end. This is supported by the argument that mild drought could enhance CO_2 uptake as it suppresses respiration but has less impact on gross primary productivity (Grant et al., 2006). These reasonable non-linear, cross-correlated, multi-dimensional ERFs were able to be extracted from BRT and then used to project turbulent exchange maps between surface and atmosphere.

3.2. Spatial-temporally resolved flux maps

Fig. 7 shows an example of the domain scaled H, LE and CO_2 flux over the predefined 20 km ×20 km target area. The white cells in the figure are areas with state-space combinations of drivers for which no extracted response relationship exists. In this particular time slice, the ERF projection covered 73%, 54%, 62% of the target area. Over the whole experiment, the coverage was extended from the original <10% (average footprint area relative to 20 km × 20 km) to $56\% \pm 13\%$, $51\% \pm 6\%$, and $51\% \pm 13\%$ for $20 \times 20\ km^2$ target domain for H, LE and CO_2 flux respectively, where the tolerance here is one standard deviation. The spatial coverage of the prediction area during the experiment time period was not constant (Fig. 8) and increased with smaller target domain area. Spatial coverage information can be used to identify the representativeness of tower observations in future study.

In this study, when summarized over the whole experiment, the projected H, LE and CO_2 flux differed from the tower observations in their expected value (+27%, −9%, and −17%) and spatio-temporal variation (−22%, −21%, and −3%). The mean of total domain projected H, LE and CO_2 flux over a 20 km × 20 km target domain were 50.1 $W\,m^{-2}$, 44.8 $W\,m^{-2}$ and −94.4 $gC\,month^{-1}\,m^{-2}$, compared to the flux observations, 39.6 $W\,m^{-2}$, 49.2 $W\,m^{-2}$, and −80.4 $gC\,month^{-1}\,m^{-2}$. One standard deviation of spatio-temporal variation of projected H, LE, and CO_2 flux was 101 $W\,m^{-2}$, 73 $W\,m^{-2}$, and 322 $gC\,month^{-1}\,m^{-2}$, smaller than observations, 124 $W\,m^{-2}$, 92 $W\,m^{-2}$, 332 $gC\,month^{-1}\,m^{-2}$, respectively. Smaller carbon uptake measured by WLEF was consistent with Desai et al. (2008), stating that WLEF under-estimated carbon sinks compared to aggregation of nearby flux towers. CO_2 flux difference is likely owing to a higher LST and EVI across the domain compared to the footprint, in particular during daytime (Fig. 2). Similar in H and LE: Larger H over target area is due to the positive relationship between LST and H in extracted ERFs, although EVI has a small opposite effect due to large EVI in target area; Domain-scaled LE is slightly smaller than observed, likely owing to its slight negative relationship with LST within the study range of 291 K–295 K (not shown because of limited space). Flux differences reflect the effect of transient footprint bias and differences in mean surface properties in the target area compared to the flux footprint. In addition, the greater domain-scale heat turbulent exchange may potentially point to a source of non-closure of energy balance (Stoy et al., 2013; Foken, 2008). Smaller spatio-temporal variation of projections compared to observations was likely the result of the smaller spatio-temporal variation of surface properties over the fixed target area than the temporally spatially varying footprint area.

3.3. Uncertainty budget

The median systematic and random uncertainty terms (median absolute deviation, in parenthesis) are shown for each uncertainty source in Table 1. The first two rows are uncertainty for a single observation, whereas rows 3–6 are uncertainty per projection.

The uncertainty from instrumentation and hardware was estimated as detection limit, below which the instrument could not

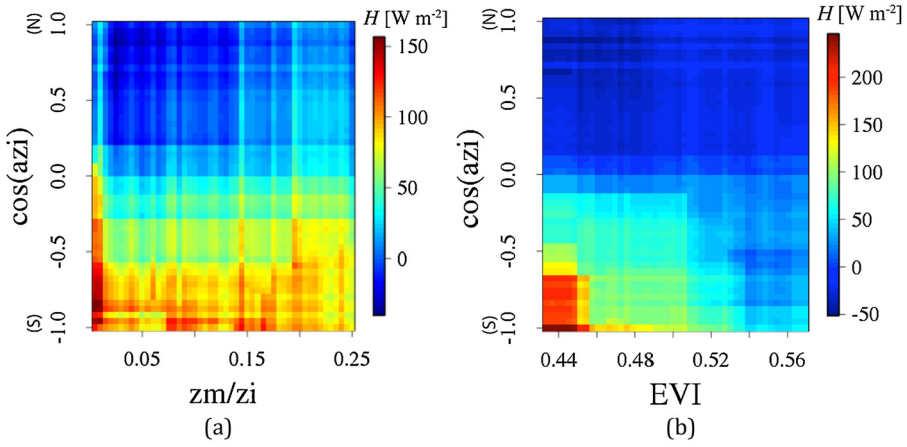

Fig. 6. The response surface of sensible heat as function of (a) cos(azi) and z_m/z_i, and (b) cos(azi) and EVI.

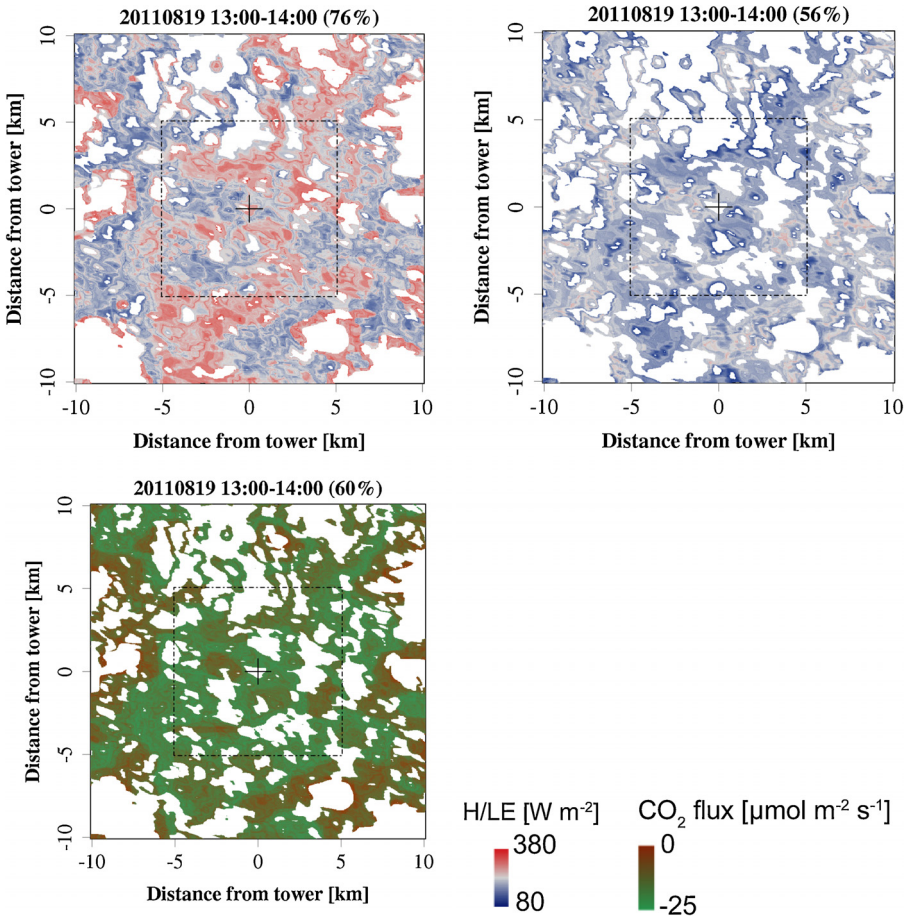

Fig. 7. Averaged minutely projected flux grids of sensible heat (top left panel), latent heat (top right panel) and CO_2 (bottom left panel) over August 19th, 2011, 13:00–14:00 CST for tower-centered 20 × 20 km^2 target region. White areas are gaps that cannot be reproduced by ERF because their biophysical properties exceed the range of the training dataset.

distinguish signal from noise. For a single observation, the detection limit was 0.9 ± 1.2 W m^{-2} for H, 1.4 ± 1.8 W m^{-2} for LE and 0.2 ± 0.3 μ mol m^{-2} s^{-1} for CO_2 flux. For H, LE and CO_2 flux, the systematic sampling errors were 1%, 1% and 0% for the Wavelet transforms of 1-hourly datasets, and the random errors were 45%, 78% and 52% for the five-minute averaging window over the Wavelet cross-scalograms, respectively.

For each single projection, the uncertainty from the unaccounted spatio-temporal variability in the state variables was 1% biased for H and unbiased for LE and CO_2 flux. The uncertainty from spatio-temporal analysis was unbiased for H and LE, and biased for CO_2 flux by −4%. The systematic uncertainty from ERF training was 0% and precise to 4% for H, 2% for LE and 2% for CO_2 flux. Finally, ERF projection uncertainty was used to assess how well the procedure performed if the footprints did not cover the predefined domain.

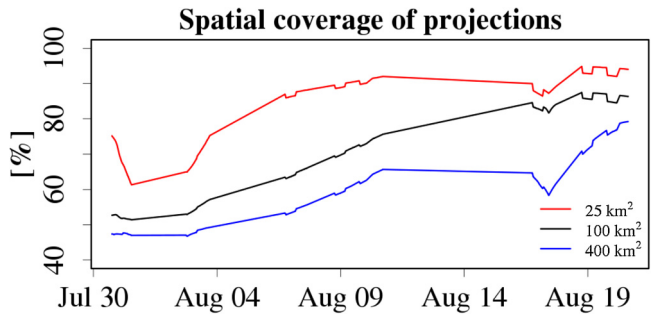

Fig. 8. Time-varying spatial coverage of the sensible heat flux projection over $5 \times 5\,km^2$ (red), $10 \times 10\,km^2$ (black) and $20\,km \times 20\,km$ area (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Here, we saw the ERF approach underestimated H by 11%, LE by 1.5% and overestimated CO_2 flux by 16%.

The ERF projection uncertainty was found to be the dominant source of systematic error for the tower ERF procedure. On this basis, an overall accuracy of <20% for H and CO_2 fluxes and <5% for LE can be assigned. Ensemble random uncertainty σ_{ens}, with $\sigma_{ens} = \frac{\sigma_{ran}}{\sqrt{N}}$, dropped rapidly when aggregating multiple observations and projection cells/periods N. Even by adding random uncertainty of all terms in quadrature, the ensemble random uncertainty for hourly projection over the whole projection domain will be confined to within 1% for all three fluxes.

3.4. Comparison with measurements and other studies

The averaged ERF-projection over 3 km × 3 km region centered by WCr explained 76% and 80% of the variation of H observed at WCr when using WCr and WLEF meteorological drivers for projection, respectively (Fig. 9). ERF-projection underestimated (slope) observations by 17% and 9% with residual standard deviation at $7.49\,W\,m^{-2}$ and $0.99\,W\,m^{-2}$. It is worth noting that the underestimation is in the same order as the systematic ERF projection error (11% in Table 1), thus corroborating the usefulness and validity of ERF-extraction and uncertainty budget of the algorithm. The difference between Fig. 9a and Fig. 9b reflects the impact of the assumption of homogeneity in meteorological drivers and it indicates that this assumption is acceptable in this study. The residual and offset could be attributed to the significant distance of the tall tower from the target tower (21 km), inability to extrapolate for all grid cells within the target WC-centered area. However, overall this independent evaluation shows agreement between the ERF-projected fluxes and the US-WCr observations.

When compared to other upscaling approaches (Table 2), our estimate, $-94 \pm 31\,gC\,m^{-2}\,month^{-1}$ is consistent with a footprint-debiased estimate $-86 \pm 18\,gC\,m^{-2}\,month^{-1}$ and a multi-tower synthesis aggregation $-97 \pm 30\,gC\,m^{-2}\,month^{-1}$ by Desai et al. (2008) over $40 \times 40\,km^2$ around WLEF in June–Aug in 2002 and 2003; and also agrees with $-11 - -107\ gC m^{-2} month^{-1}$ regional flux estimates (Wang et al., 2006) focusing on May-Sep in 2003, although over different time period and space. The uncertainty here and for those reported below represent one sd of spatial variation. However, our estimate suggests a larger carbon sequestration than Xiao et al. (2014) estimate ($-73 \pm 28 gC m^{-2} month^{-1}$) over the same area during the same period. When lowering our domain-scaled projection by 17.4% for location bias and 33.4% for combined location bias and systematic ERF projection uncertainty, we obtain $80\,gC\,m^{-2}\,month^{-1}$ and $62\,gC\,m^{-2}\,month^{-1}$, respectively. Xiao et al. (2014) estimate falls well into this range of the CO_2 location bias, indicating how current upscaling procedures might be directly impacted by the effects of scale discrepancy, such as through the neglect of changing flux footprints: Xiao et al. (2014) trained the model with tower-surrounded $3 \times 3\,km^2$ domain drivers, and the extracted relationships could be distorted due to neglecting of footprint in drivers. Therefore, one advantage of ERF method is that ERF-scaled flux grids are reliable and uncertainty directly estimated because the extracted ERFs are trained with fluxes and footprint-weighted drivers at high temporal and spatial resolution.

Among other upscaling approaches, the other advantage of ERF method is that it produces continuous hourly flux maps by involving temporally resolved (i.e. sub-hourly) meteorological variables. Direct variation of H, LE and CO_2 flux can be captured at the hourly interval and the spatiotemporal variability of projected maps can be quite different than projection at coarse temporal resolution from continental upscaling (Xiao et al., 2014, Table 2). This could exert large influence on model-data comparison as most ESMs are run at fine time resolutions and model formulations for surface fluxes were derived for this timescale of response, e.g. light use efficiency model. Further, only at hourly or finer temporal resolution can the problem of removing the footprint diurnal cycle from the true regional flux diurnal cycle of surface be accomplished.

3.5. Prospect for model-data fusion

ERF method can improve mechanistic models in two potential ways: providing model process-structures reference and by directly evaluating model fluxes against ERF-rectified flux maps at the same space and time of model grids. ERF itself essentially acts as a transfer function that can be used for parameterizing formulations or testing competing model processes. For example, a parameterization can be fitted to the response surface between H, cos(azi) and EVI and compared with empirical models. Also, if extracted ERFs are non-linear, it could be used as an indicator to suggest models should operate at higher spatial or temporal resolution: In this study, one-month integrated H over target domain was 45% greater than projected H from one-month integrated drivers.

The ERF method can also improve model evaluation by providing hourly PDFs and descriptive statistics from domain-scaled fluxes as reference: Over the study period the domain-scaled fluxes of H, LE and CO_2 were 27% larger, 9% smaller and 17% smaller than the tower observations. Here in Fig. 10, we also show that at hourly scale, PDFs of projected flux grids behaved differently from observations due to the transient footprint bias. In this case, sometimes the tower sourced areas with near-identical flux (Fig. 10a, c), displaying smaller variation than the region of a model cell. At other times, the tower might see areas exhibiting different source strength, making measurements exhibit larger variation (Fig. 10b). Therefore, ERF-rectified flux grids mitigate the pitfall of limited and varying spatial representativeness when using observations to inform a model during execution.

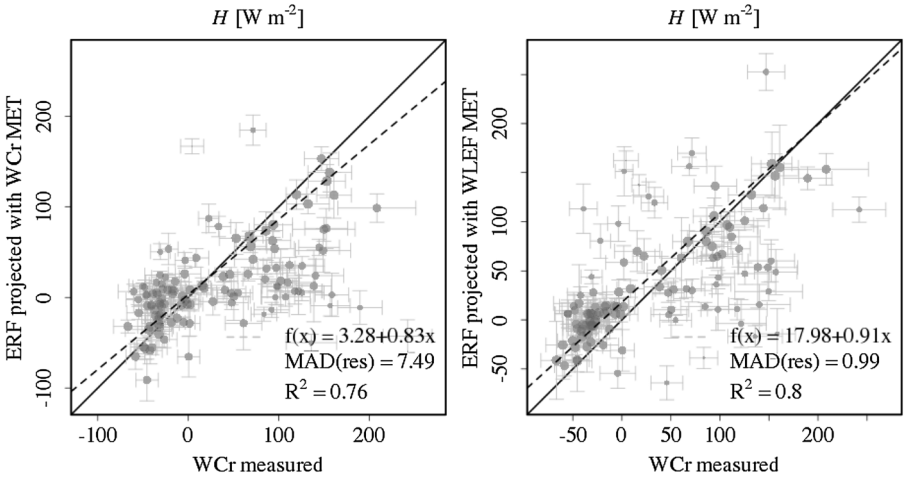

Fig. 9. Scatterplot for Willow Creek (WCr) sensible heat flux measurements and ERF projections over 3 km × 3 km centered around WCr. Each point represents a one hour averaging period, and corresponding aggregation is applied to the minutely ERF projection with meteorological drivers measured at WCr (panel a) and WLEF (panel b), respectively. Uncertainties are one standard deviation of random uncertainty and shown as error bars in x and y direction.

Table 2
ERF-projected CO_2 net ecosystem exchange [gC m^{-2} $month^{-1}$] for major land covers in the 20 km × 20 km target domain compared to other scaling approaches applied around WLEF. Where available, either the range of values or one standard deviation of spatial variation is provided.

	ERF-projection	Xiao et al., 2014	Desai et al., 2008 (footprint decomposition)	Wang et al., 2006	Desai et al., 2008 (Multi-tower synthesis aggregation)
Target domain	−94 ± 31	−73 ± 28	−86 ± 19	−11 to −107	−97 ± 30
Woody wetland	−91 ± 31	−71 ± 21	−66	−46	–
Deciduous forest	−98 ± 31	−72 ± 22	−132	–	–
Mixed forest	−96 ± 35	−70 ± 23	–	−49	–

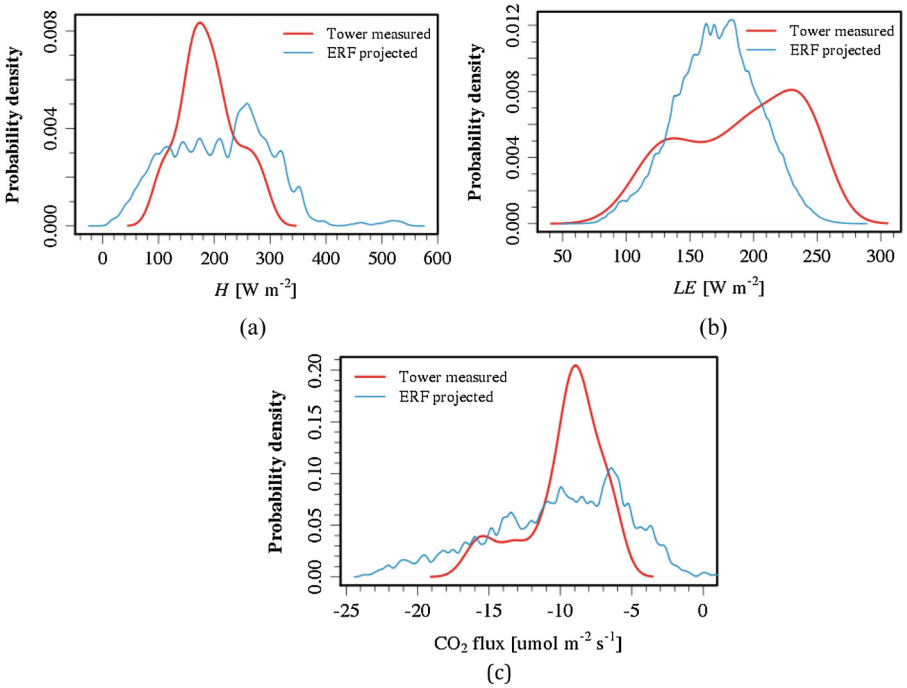

Fig. 10. Probability density functions of sensible heat (*H*, panel a), latent heat (LE, panel b) and CO_2 (panel c) turbulent fluxes. Shown are minutely tower observations (red) and minutely ERF flux maps over 20 × 20 km^2 (blue) for August 19th, 2011, 13:00 14:00 CST. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.6. Current limitations and future directions

Some factors or assumptions limit and degrade current tower ERF performance, which can be further researched in future studies focusing on improving the ERF approach.

First, at sub-hourly time scales, surface-atmosphere exchange is rarely resolved completely by the turbulent flux alone. Specifically in the case of taller towers, storage beneath the turbulent

flux measurement height can comprise a substantial amount of the actual surface-atmosphere exchange. Through combining storage and turbulent flux, an improved performance in mapping net ecosystem exchange is expected, which bears the potential to address surface-heterogeneity-related components of a frequently observed non-closed energy balance.

Second, an inappropriate assumption in ERF projection is the homogeneity of the meteorological drivers across the target domain. Further study is undergoing focusing on the impact of spatial heterogeneity of meteorological states on the performance of ERF approach using gridded reanalysis meteorology with observations. Even with spatially resolved drivers, the spatial resolution can have a large influence on ERF's power to detect the surface spatial variability. The resulting uncertainty is influenced by how many pixels are sampled by towers, especially for short towers. Potential high-resolution (both spatial and temporal) remote sensing products provide an opportunity to improve the representation of small-scale spatiotemporal heterogeneity in ERFs.

Third, the projection spatial coverage could not reach 100%, as this method chooses to only project to surfaces with properties that have been experienced by machine learning through training data. Alternate machine learning approaches or interpolation methods may need to be implemented if the ultimate goal is a full resolved projected grid.

It is also worthy to note that when applying BRT at hourly or finer resolution, drivers should include one with strong diurnal cycle, e.g. azimuth angle. This resolved issues of the diurnal cycle in footprint variation resulting in biophysically implausible response attribution across drivers. In addition, here we tested ERF on a site with known heterogeneity and suspected footprint biases arising from it. Whether more homogeneous sites can also benefit from the application of ERF is an area of more work. However, in many cases even ostensibly homogenous sites can exhibit large spatial variation in fluxes, especially for trace gases like methane. Here, ERF could help disentangle processes and patterns, provided suitable land surface drivers are available.

4. Conclusions

Our study demonstrated applicability of the environmental response function (ERF) approach to map heat and CO_2 fluxes from an eddy-covariance tower to the scale of an earth system model grid cell. In comparison to other upscaling methodologies, the two main advantages of the ERF approach are the explicit consideration of varying flux footprints during training, and the ability to produce regional, high-resolution flux grids at hourly timescales.

ERF-projected flux grids not only assess, but also rectify the spatial representativeness of tower eddy-covariance measurements. This is achieved through decomposing and spatio-temporally attributing heterogeneous surface contribution. It thus provides a promising tool for studying heterogeneity-induced non-closure of the surface energy balance, and for interpreting flux observations in model-data fusion approaches. For example, mechanistic models can be improved through ERF-extracted process-structures, and can be constrained using ERF-projected hourly flux and uncertainty maps.

Applying the ERF procedure to longer time periods and across multiple tower sites can help decomposing the interplay of seasonal, inter-annual and spatial variability on regional scales. The main limitations of ERF approach in this study were the omission of storage flux, the assumption of spatially homogeneous meteorological drivers during projection, and state-space gaps in the projected grids, which are areas of ongoing and future research.

Acknowledgements

This work was supported by NEON, Inc. contact #1010-0196-000 to U. Wisconsin, DOE Office of Science Ameriflux Network Management Project subaward to ChEAS Core Site Cluster, and NSF Advances in Biological Informatics awards #1062204 and #1457897.

References

Ammann, C., Brunner, A., Spirig, C., Neftel, A., 2006. Technical note: water vapour concentration and flux measurements with PTR-MS. Atmos. Chem. Phys. 6 (12), 4643–4651, http://dx.doi.org/10.5194/acp-6-4643-2006.

Arora, V.K., Boer, G.J., Friedlingstein, P., Eby, M., Jones, C.D., Christian, J.R., Cadule, P., 2013. Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models. J. Clim. 26 (15), 5289–5314, http://dx.doi.org/10.1175/JCLI-D-12-00494.1.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Evans, R., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82 (11), 2415–2434, http://dx.doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3. CO;2.

Billesbach, D.P., 2011. Estimating uncertainties in individual eddy covariance flux measurements: a comparison of methods and a proposed new method. Agric. For. Meteorol. 151 (3), 394–405, http://dx.doi.org/10.1016/j.agrformet.2010.12.001.

Bonan, G.B., Lawrence, P.J., Oleson, K.W., Levis, S., Jung, M., Reichstein, M., Swenson, 2011. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. J. Geophys. Res.: Biogeosci. 116 (G2), 2005–2012, http://dx.doi.org/10.1029/2010jg001593.

Brock, F.V., 1986. A nonlinear filter to remove impulse noise from meteorological data. J. Atmos. Oceanic Technol. 3 (1), 51–58, http://dx.doi.org/10.1175/1520-0426(1986)003<0051:ANFTRI>2.0. CO;2.

Budishchev, A., Mi, Y., van Huissteden, J., Belelli-Marchesini, L., Schaepman-Strub, G., Parmentier, F.J.W., Dolman, A.J., 2014. Evaluation of a plot-scale methane emission model using eddy covariance observations and footprint modelling. Biogeosciences 11 (17), 4651–4664, http://dx.doi.org/10.5194/bg-11-4651-2014.

Businger, J.A., Wyngaard, J.C., Izumi, Y., Bradley, E.F., 1971. Flux-Profile relationships in the atmospheric surface layer. J. Atmos. Sci. 28 (2), 181–189, http://dx.doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0. CO;2.

Cacuci, D.G., 2003. Sensitivity & Uncertainty Analysis Volume 1: Theory. CRC Press.

Collins, W.J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Woodward, S., 2011. Development and evaluation of an Earth-System model HadGEM2. Geosci. Model Dev. 4 (4), 1051–1075, http://dx.doi.org/10.5194/gmd-4-1051-2011.

Cook, B.D., Davis, K.J., Wang, W., Desai, A.R., Berger, B.W., Teclaw, R.M., Martin, J.G., Bolstad, P.V., Bakwin, P.S., Yi, C., Heilman, W., 2004. Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA. Agric. Forest Meteorol. 126 (3-4), 271–295, http://dx.doi.org/10.1016/j.agrformet.2004.06.008.

Croux, C., Rousseeuw, P.J., 1992. Time-efficient Algorithms for Two Highly Robust Estimators of Scale. Springer.

Desai, A.R., Noormets, A., Bolstad, P.V., Chen, J., Cook, B.D., Davis, K.J., Ricciuto, D.M., 2008. Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: implications for regional scaling. Agric. Forest Meteorol. 148 (2), 288–308, http://dx.doi.org/10.1016/j.agrformet.2007.08.001.

Desai, A.R., Helliker, B.R., Moorcroft, P.R., Andrews, A.E., Berry, J.A., 2010. Climatic controls of interannual variability in regional carbon fluxes from top-down and bottom-up perspectives. J. Geophys. Res.: Biogeosci. 2005–2012 (115(G2)), http://dx.doi.org/10.1029/2009JG001122.

Desai, A.R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., Kolka, R., 2015. Landscape-level terrestrial methane flux observed from a very tall tower. Agric. Forest Meteorol. 201, 61–75, http://dx.doi.org/10.1016/j.agrformet.2014.10.017.

Desai, A.R., 2014. Influence and predictive capacity of climate anomalies on daily to decadal extremes in canopy photosynthesis. Photosynth. Res. 119 (1–2), 31–47, http://dx.doi.org/10.1007/s11120-013-9925-z.

Dietze, M.C., Serbin, S.P., Davidson, C., Desai, A.R., Feng, X., Kelly, R., McHenry, K., 2014. A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. J. Geophys. Res.: Biogeosci. 119 (3), 286–300, http://dx.doi.org/10.1002/2013JG002392.

Dufresne, J.L., Foujols, M.A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Vuichard, N., 2013. Climate change projections using the IPSL-CM5 earth system model: from CMIP3 to CMIP5. Clim. Dyn. 40 (9–10), 2123–2165, http://dx.doi.org/10.1007/s00382-012-1636-1.

Foken, T., Nappo, C.J., 2008. Micrometeorology. Springer.

Foken, T., 2008. The energy balance closure problem: an overview. Ecol. Appl. 18 (6), 1351–1367, http://dx.doi.org/10.1890/06-0922.1.

Grant, R., Black, T., Gaumont-Guay, D., Klujn, N., Barr, A., Morgenstern, K., Nesic, Z., 2006. Net ecosystem productivity of boreal aspen forests under drought and

climate change: mathematical modelling with Ecosys. Agric. Forest Meteorol. 140 (1), 152–170, http://dx.doi.org/10.1016/j.agrformet.2006.01.012.
Högström, U., 1988. Non-dimensional wind and temperature profiles in the atmospheric surface layer: a re-evaluation. In: Topics in Micrometeorology. A Festschrift for Arch Dyer. Springer, pp. 55–78.
Homer, C., Dewitz, J., Yang, L.M., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous United States - representing a decade of land cover change information. Photogramm. Eng. Remote Sens. 81 (5), 345–354, http://dx.doi.org/10.14358/pers.81.5.345.
Hutjes, R., Vellinga, O., Gioli, B., Miglietta, F., 2010. Dis-aggregation of airborne flux measurements using footprint analysis. Agric. Forest Meteorol. 150 (7), 966–983, http://dx.doi.org/10.1016/j.agrformet.2010.03.004.
Jacinthe, P.-A., Lal, R., 2006. Spatial variability of soil properties and trace gas fluxes in reclaimed mine land of southeastern Ohio. Geoderma 136 (3), 598–608, http://dx.doi.org/10.1016/j.geoderma.2006.04.020.
Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., De Jeu, M.L., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467 (7318), 951–954, http://dx.doi.org/10.1038/nature09396.
Kljun, N., Calanca, P., Rotach, M., Schmid, H., 2004. A simple parameterisation for flux footprint predictions. Boundary Layer Meteorol. 112 (3), 503–523, http://dx.doi.org/10.1023/B:BOUN.0000030653.71031.96.
Le Quéré, C., Peters, G., Andres, R., Andrew, R., Boden, T., Ciais, P., Moriarty, R., 2013a. Global carbon budget 2013. Earth Syst. Sci. Data Discuss 6 (2), 689–760, http://dx.doi.org/10.5194/essdd-6-689-2013.
Le Quéré, C., Andres, R.J., Boden, T., Conway, T., Houghton, R., House, J.I., Ahlström, A., 2013b. The global carbon budget 1959–2011. Earth Syst. Sci. Data 5 (1), 165–185, http://dx.doi.org/10.5194/essd-5-165-2013.
Le Quéré, C., Moriarty, R., Andrew, R., Peters, G., Ciais, P., Friedlingstein, P., Arneth, A., 2014. Global carbon budget 2014. Earth Syst. Sci. Data Discuss. 7 (2), 521–610, http://dx.doi.org/10.5194/essdd-7-521-2014.
Lenschow, D.H., Stankov, B.B., 1986. Length scales in the convective boundary layer. J. Atmos. Sci. 43 (12), 1198–1209, http://dx.doi.org/10.1007/s10546-004-2165-1.
Lenschow, D., Mann, J., Kristensen, L., 1994. How long is long enough when measuring fluxes and other turbulence statistics? J. Atmos. Oceanic Technol. 11 (3), 661–673, http://dx.doi.org/10.1175/1520-0426(1994)011<0661:hlilew>2.0. co;2.
Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H.P., Foken, T., 2012. Eddy-covariance flux measurements with a weight-shift microlight aircraft. Atmos. Meas. Tech. 5 (7), 1699–1717, http://dx.doi.org/10.5194/amt-5-1699-2012.
Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón y Widemann, B., Neidl, F., Foken, T., 2013. Spatially explicit regionalization of airborne flux measurements using environmental response functions. Biogeosciences 10 (4), 2193–2217, http://dx.doi.org/10.5194/bg-10-2193-2013.
Papale, D., Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Global Change Biol. 9 (4), 525–535, http://dx.doi.org/10.1046/j.1365-2486.2003.00609.x.
Prueger, J., Alfieri, J., Hipps, L., Kustas, W., Chavez, J., Evett, S., McKee, L., 2012. Patch scale turbulence over dryland and irrigated surfaces in a semi-arid landscape under advective conditions during BEAREX08. Adv. Water Resour. 50, 106–119, http://dx.doi.org/10.1016/j.advwatres.2012.07.014.
Development Core Team, R., 2012. R: A language and environment forstatistical computing. R foundation for statistical computing, Vienna, Austria.
Ripley, B.D., Thompson, M., 1987. Regression techniques forthe detection of analytical bias. Analyst 112, 377–383, http://dx.doi.org/10.1039/an9871200377.
Richardson, A.D., Anderson, R.S., Arain, M.A., Barr, A.G., Bohrer, G., Chen, G., Desai, A.R., 2012. Terrestrial biosphere models need better representation of vegetation phenology: results from the North American carbon program site synthesis. Global Change Biol. 18 (2), 566–584, http://dx.doi.org/10.1111/j.1365-2486.2011.02562.x.
Rousseeuw, P.J., Verboven, S., 2002. Robust estimation in very small samples. Comput. Stat. Data Anal. 40 (4), 741–758, http://dx.doi.org/10.1016/S0167-9473(02)00078-6.
Schaefer, K., Schwalm, C.R., Williams, C., Arain, M.A., Barr, A., Chen, J.M., Hollinger, D.Y., 2012. A model-data comparison of gross primary productivity: results from the North American carbon program site synthesis. J. Geophys. Res.: Biogeosci. 117 (G3), http://dx.doi.org/10.1029/2012jg001960.
Schotanus, P., Nieuwstadt, F., De Bruin, H., 1983. Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes. Bound. Layer Meteorol. 26 (1), 81–93, http://dx.doi.org/10.1007/BF00164332.
Schwalm, C.R., Williams, C.A., Schaefer, K., Anderson, R., Arain, M.A., Baker, I., Chen, J.M., 2010. A model–data intercomparison of CO2 exchange across north america: results from the North American carbon program site synthesis. J. Geophys. Res.: Biogeosci. 115 (G3), http://dx.doi.org/10.1029/2009JG001229.
Starkenburg, D., Fochesatto, G.J., Cristóbal, J., Prakash, A., Gens, R., Alfieri, J.G., Kane, D.L., 2015. Temperature regimes and turbulent heat fluxes across a heterogeneous canopy in an Alaskan boreal forest. J. Geophys. Res.: Atmos. 120 (4), 1348–1360, http://dx.doi.org/10.1002/2014JD022338.
Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Bernhofer, C., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: the role of landscape scale heterogeneity. Agric. For. Meteorol. 171, 137–152, http://dx.doi.org/10.1016/j.agrformet.2012.11.004.
Strunin, M., Hiyama, T., 2004. Applying wavelet transforms to analyse aircraft-measured turbulence and turbulent fluxes in the atmospheric boundary layer over eastern Siberia. Hydrol. processes 18 (16), 3081–3098.
Sulkava, M., Luyssaert, S., Zaehle, S., Papale, D., 2011. Assessing and improving the representativeness of monitoring networks: the European flux tower network example. J. Geophys. Res.: Biogeosci. 116 (G3), http://dx.doi.org/10.1029/2010jg001562.
Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79 (1), 61–78, http://dx.doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.
Wang, W., Davis, K.J., Cook, B.D., Butler, M.P., Ricciuto, D.M., 2006. Decomposing CO2 fluxes measured over a mixed ecosystem at a tall tower and extending to a region: a case study. J. Geophys. Res.: Biogeosci. 111 (G2), http://dx.doi.org/10.1029/2005JG000093.
Wilczak, J.M., Oncley, S.P., Stage, S.A., 2001. Sonic anemometer tilt correction algorithms. Bound. Layer Meteorol. 99 (1), 127–150, http://dx.doi.org/10.1023/A.;1;1018966204465.
Williams, M., Richardson, A., Reichstein, M., Stoy, P., Peylin, P., Verbeeck, H., Kattge, J., 2009. Improving land surface models with FLUXNET data. Biogeosciences 6 (7), 1341–1359, http://dx.doi.org/10.5194/bg-6-1341-2009.
Xiao, J., Davis, K.J., Urban, N.M., Keller, K., Saliendra, N.Z., 2011. Upscaling carbon fluxes from towers to the regional scale: influence of parameter variability and land cover representation on regional flux estimates. J. Geophys. Res. 116 (G3), http://dx.doi.org/10.1029/2010jg001568.
Xiao, J., Chen, J., Davis, K.J., Reichstein, M., 2012. Advances in upscaling of eddy covariance measurements of carbon and water fluxes. J. Geophys. Res.: Biogeosci. 117 (G1), http://dx.doi.org/10.1029/2011jg001889.
Xiao, J., Ollinger, S.V., BFrolking, S., Hurtt, G.C., Hollinger, D.Y., Davis, K.J., Suyker, A.E., 2014. Data-driven diagnostics of terrestrial carbon dynamics over North America. Agric. For. Meteorol. 197, 142–157, http://dx.doi.org/10.1016/j.agrformet.2014.06.013.
Zhou, Z.-H., 2012. Ensemble Methods: Foundations and Algorithms. CRC Press.