Scaling of the Carbon, Energy Exchange, and Evapotranspiration over Heterogeneous Ecosystems

by

Ke Xu

A dissertation for the Degree of Doctor of Philosophy

(Atmospheric and Oceanic Sciences)

at the

University of Wisconsin-Madison

March 16, 2018

The dissertation is reviewed and approved* by the following members of the Final Oral Committee:

Ankur Desai,

Stefan Metzger,

Tristan L'Ecuyer,

Dan Vimont,

Chris Kucharik,

*Signatures are on file in the Graduate School

Acknowledgements

I am gratefully indebted to my advisor and friend, Prof. Ankur R. Desai, who has always encouraged and cared about me, instead of pushed me. This really helps me deal with my procrastination especially in this PhD research. Equally important has been the flexible accessibility, almost immediate response to emails, insightful thoughts, sharp comments, suggestions on research and kid raising, and financial support through his grant. Like many students in the lab, I am constantly reminded of how lucky I was to have such a great advisor. Great thanks also go to my mentor, Dr. Stefan Metzger, who is not only an expert but also a friend to me. Stefan has got me onboard on the wonderful Environmental Response Function (ERF) method, invited me to intern at Surface-Atmosphere Exchange team in National Ecological Observation Network (NEON) not only once but twice, and made me a NEON postdoc position after this defense. Stefan not only gives me numerous comments on my work, but also helps me to learn organize and prioritize my research in a better and more beautiful way. Special thanks also go to my committee, Profs. Tristan L'Ecuyer, Dan Vimont, and Chris Kucharik, for being flexible on all my committee meeting time, for reading this dissertation, and for their helpful suggestions, comments, and contributions.

I would like to thank all co-authors, Stefan Metzger, Ankur Desai, Matthias Sühring, and David Durden for their prompt feedbacks, revisions, and grants, on the following three papers which derived into Chapter 2–4:

Xu, K., Metzger, S., & Desai, A. R. (2017). Upscaling tower-observed turbulent exchange at fine spatio-temporalresolution using environmental response functions. Agricultural and Forest Meteorology, 232, 10-22. doi:10.1016/j.agrformet.2016.07.019

Xu, K., Metzger, S., & Desai, A. R. (2017). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. Agricultural and Forest Meteorology. doi:10.1016/j.agrformet.2017.10.011

Xu, K., Sühring, M., Metzger, S., Durden, D., & Desai, A. R. Can data mining help eddy-covariance see the landscape? A large-eddy simulation study, In Preparation

I would also like to thank all members in Desai's lab for all their encouragement, help, and discussion in the lab, including Jonathan Thom, Gosia Golub, Sean DuBios, Tommy Jasmin, David Reed, James Simkins, Ammara Talib, Molly Aufforth, Ryan Clare, Lewis Kunik, Gabe Bromley, Zach Hansen, Ben Sulman, Manuel Helbig. Especial thanks go to our lab manager, Jonathan Thom, who is dedicated to ensure 200+ instruments functioning on all our 7 tower sites. I also appreciate the entire Department of Atmospheric and Oceanic Sciences, all teachers giving multiple great courses, and the wonderfully-organized colloquium, CPEP seminars, and department seminars. I am grateful to Profs. Tristan L'Ecuyer, Ankur Desai, Chris Kucharik, Grant Petty, Matt Hitchman, Zhengyu Liu, and Galen Mckinlay.

The most difficult part of this PhD work was done during my internships at NEON, and I would like to thank all scientists and friends I met at NEON for their contribution to data products, model output, and deep discussion to make this dissertation possible. My warm thanks go to Stefan Metzger, Hongyan Luo, Natchaya Durden, David Durden, Rommel Zulueta, and Cove Sturtevant.

It was the first research experience working with Prof. Weimin Ju at the Nanjing University, China, that intrigued me to the scientific research. I would gratefully thank the professors, lab members and friends in Prof. Ju's huge lab: Yibo Liu, Yanlian Zhou, Xiaocui Wu, Mingzhu He, Guang Zheng, Qian Zhang, Chunhua Zhang, Sheng Wang, Qin Ma, Yang Ju, Bailin Xing, Jiande Ma, Weiliang Fan, Dengqiu Li, Mei Zan, Jinlong Huang.

Numerous other colleagues and collaborators were instrumental in my research and I also wish to acknowledge the help involved in data production, data collection, and future expansion for the work in this dissertation. Those include: Andy Fox, David Lawrence, Martyn Clark, NCAR, Mike Dietze, Boston University, Chosen Chu, Dennis Baldocchi, Trevor Keenan, Lawrence Berkeley National Lab, Josh Fisher, NASA JPL, Sebastain Wolf, ETH Zurich, Róisín Commane, Harvard University and Columbia University, Gretchen Keppel-Aleks, U of Michigan, Ken Davis, Pennsylvania State University, Matthias Mauder, KIT, Dave Moore, U of Arizona.

My friends in my graduate program also deserve credit for the helpful discussion and friendship. I am appreciated of Yan Yu, Yanghui Kang, Haidi Chen, April Hang, Huaran Liu, Fuyao Wang, Shu Wu, Hanzhi Zhang, Jun Hu, Feng He, Pei Wang, Sifan Gu, Jiaxu Zhang, and Kai-yuan Zheng.

I would like to thank all the funding supporting this work, 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, and the National Science Foundation grant #DBI-0752017 and #DEB-0845166, and travel funding from both U W-Madison and AOS department.

Last, I have greatest debt to my dedicated mama and baba, and my husband, Jiang Zhu, whom I met in this department, for their continuous support and patience. This dissertation is dedicated to my husband and our first kid, Victor Zhu, whose curiosity to the world, optimistic, and infinite reservoir of hugs never failed to counteract the stresses coming with getting a Ph.D.

Abstract

Surface-atmosphere exchanges of energy, water vapor, and chemical materials are key indicator climate variables for studying the impact of the terrestrial biosphere response to changes in climate and land-surface feedbacks. For decades, eddy-covariance observations have been used extensively to monitor these surface-atmosphere exchange and to better understand ecosystem functioning, climate impact, and the carbon cycle.

However, two challenges, the location bias and energy imbalance, prevent eddy-covariance data from accurately sampling the regional scale of land surface processes and the corresponding atmospheric responses. These two challenges have limited our ability to better benchmark Earth system models and land surface remote sensing with eddy-covariance data, and to better constrain parameterizations for modeling the atmosphere, landscape and urban ecology. Previous studies suggested these two challenges can be mitigated by applying upscaling methods and by adding multiple towers over the regional domain.

In this dissertation, I ask three key questions: Is it possible to rectify location bias, by extending the application of the environmental response function (ERF) upscaling method, from the airborne eddy-covariance to tower eddy-covariance measurements? Can ERF close energy budget in eddy-covariance observations by resolving the not-easy-measured storage flux and vertical advection flux? How does ERF perform compared to previous scaling approaches using one or multiple towers?

My dissertation has developed and evaluated a novel upscaling approach, ERF, for tower-based eddy-covariance observations. The method enables mapping surface-atmosphere exchanges to the regional scale by utilizing tower flux footprint variation over heterogeneous surfaces at high spatiotemporal resolution. In an application using a real tower site over a heterogeneous surface, I found ERF enables improving sensible and latent heat flux low bias by 21% by

detecting and resolving mesoscale eddies and storage flux. Compared to previous upscaling methods and adding more towers, ERF has advantages in reducing observation requirements by 80%, and in enabling flux decomposition into component fluxes at high spatio-temporal resolution. As a result, ERF resolves long-standing footprint bias and energy imbalance problems in observing net surface-atmosphere exchange from eddy-covariance flux towers. These results have significant applications for advancing model-data fusion, remote sensing applications, and boundary layer studies.

Contents

1 Introduction

In this chapter, I provide an overview of the motivation, challenges, and relevant literature for this dissertation (Sect. 1.1). The background of eddy-covariance technique is described in Sect. 1.2. Lastly, I give an overview of the whole dissertation (Sect. 1.3).

1.1 Motivation

How terrestrial biosphere responds to climate changes is one of the most important scientific and societal questions facing humanity. The uncertainty of this question remains one of the largest sources of uncertainty in climate projections because ecosystems can act as either carbon sinks (photosynthesis) or carbon sources (respiration, decomposition, mortality, combustion), and provide climate feedbacks through latent heat fluxes, albedo, surface roughness, and water cycling (Friedlingstein et al., 2014). Given recent and projected climate changes, the need to monitor and understand terrestrial response continues to increase from the perspective of the food and water security (Suni et al., 2015).

As the linkage of energy cycle, water cycle, and carbon cycle between the atmosphere and ecosystems, sensible heat, latent heat and carbon fluxes serve as the central hub to study the impact of climate on ecosystems and land-atmosphere feedbacks. Sensible heat, latent heat and carbon fluxes have been directly, continuously, and extensively monitored by the eddy-covariance technique across contrasting eco-climate regions (Metzger, 2017). Now eddy-covariance observations, such as those collected by AmeriFlux and National Ecological Observatory Network (NEON), have become available at unprecedented temporal duration and distributed spatial extents (Novick, et al., 2018). Near continuous (hourly/half-hourly) data on exchanges of carbon, water, heat and momentum are collected across North America and many

locations (Metzger, 2017). The longest running towers are now approaching three decades of observations (Baldocchi, 2008). This drives eddy-covariance to be one of the most important dataset to sample land surface at regional scale and study land-atmosphere interactions, e.g. detection of localized CO_2 sources (Schwandner et al., 2017), ecosystem sensitivity to temperature, dryness, and change of phenology (Wolf et al., 2017; Desai, 2014; Mahecha et al., 2017; Tang et al., 2014).

However, eddy-covariance data have been prevented from accurately sampling the regional land surface domain by two challenges, location bias (Sect. 1.1.1) and energy balance non-closure (Sect. 1.1.2).

1.1.1 Challenge I: The location bias

The first challenge stems from footprint bias. This footprint, the surface that influences a flux measurement over an ensemble Reynolds' averaged flux, not only varies rapidly in time, but is also limited within 10^{-1}–10^{1} km^2 around the tower, which is two to three orders smaller than the regional scale relevant to landscale ecology or Earth system model grids (Figure 1). Real ecosystems are heterogeneous in many respects and on many scales. Thus, the flux footprint cannot represent the actual control volume mass balance. The flux tower footprint can be also quite variable on daily (Xu et al., 2017a), seasonal (Kim et al., 2018), and inter-annual timescales, leading to the situation where one surface type is likely observed more frequently at one time than another (Metzger, 2017). Hence, any observed dynamics or potential trends may be quite different only because of sampling a different source area (e.g., Griebel et al., 2016; Montaldo and Oren, 2016; Morin et al., 2017; Metzger 2017). Moreover, not only turbulent flux, each vertical profile observation used to calculate storage flux has its own footprint, resulting in an influence area that increases with measurement height. This can be another source of location bias in flux tower observations especially at hourly resolution

(Schmid 1997; Raupach et al., 1988; Xu et al., 2017a). As a practical workaround, flux observations are often filtered e.g. by stationarity and turbulent mixing tests for periods that more likely fulfill these assumptions.

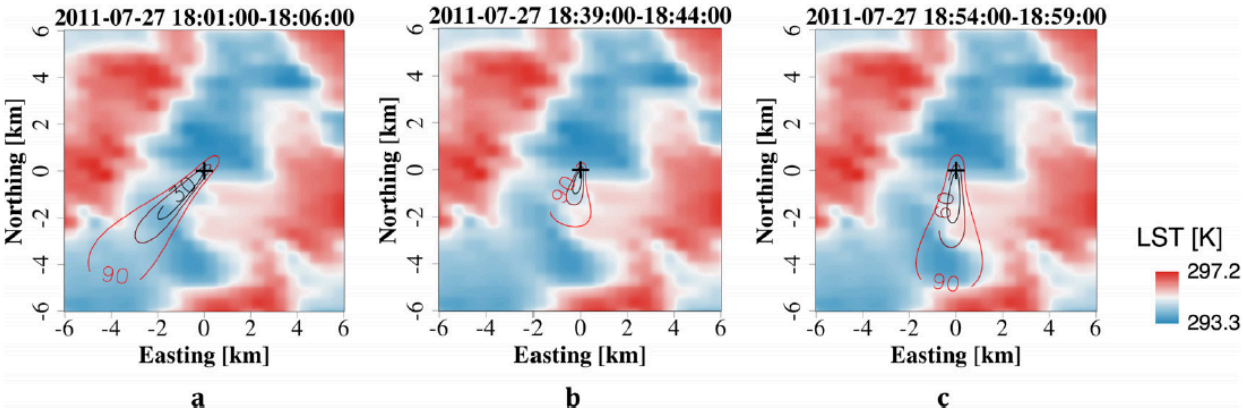


Figure 1. Fast and large variations in flux footprint with time at the AmeriFlux Park Falls tall tower at 122 m measurement height during 2011 July 27th (a) 18:01–18:06, (b) 18:39–18:44, (c) 18:54–18:59, superimposed over smoothed MODIS land surface temperature over 25×25 km^2, demonstrating the location bias between the individual tower measurement and targeted regional scale, 25×25 km^2 (Xu et al., 2017a).

Beyond data filtering, to address to address the location bias, 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, 2008; 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, 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.

More recently, environmental observations have become available at unprecedented spatial, temporal and spectral resolutions (e.g., Damian *et al*., 2014; Kampe *et al*., 2011; Musinsky *et al*. 2013). While enabling new insights into ecosystem functioning, such data also requires re-thinking scaling methodologies (e.g., Antonarakis *et al*., 2014; Hilker *et al*., 2013; Kobayashi *et al*., 2012). 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). Thus far, the ERF approach has been developed with and utilized for aircraft-based eddy-covariance measurements in the spatial domain (Metzger et al., 2013). In Chapter 2, we develop and test an ERF approach for tower-based flux observations to solve this long-standing location bias.

1.1.2 Challenge II:Energy Balance Non-closure

Scaling for flux towers is further complicated by the surface energy balance closure problem. Energy balance closure refers to an observed pattern that, at almost all sites, the observed turbulent sensible and latent heat fluxes are always 10–30% less than the sum of incoming available energy (net shortwave and long wave radiation minus ground heating), when expected to be zero at long term scale, e.g. multiple-years (Foken *et al*., 2011).

$$R = \frac{H + LE}{Q_s - Q_G} = 0.7 \sim 0.9 \qquad (3)$$

where R is energy balance ratio, which is the sum of sensible heat flux H and latent heat flux LE divided by the net radiation Q_S minus the ground heat flux Q_G.

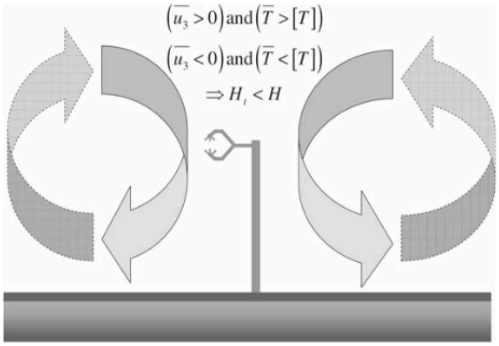


Figure 2. from Mauder *et al* (2008), a hypothesized source of "missing energy" arises from land surface heterogeneity developing stationary mesoscale eddies in a convective boundary layer.

One main reason is that low-frequency mesoscale flux contributions are hidden in the advection term (Eqn. 1, term VII), and inherently not captured by eddy-covariance technique as suggested and agreed by an increasing number of analyses (e.g. Finnigan, 2003; Kanda et al., 2004; Foken 2008; Eder et al., 2015). Atmospheric structures, i.e., turbulent organized structures (Kanda et al., 2004; Finnigan 2003) and/or secondary circulation associated with surface heterogeneity (Figure 1, Schlegel et al., 2014; Eder et al., 2015), can cause low-frequency mesoscale flux contributions. Turbulent organized structures are thermal-driven open cells (Eder et al, 2015; Wilczak and Tillman, 1980) or, in case of background wind, horizontal rolls (Drobinski et al., 1998; Maronga and Raasch, 2013) and typically have timescales larger than the averaging time of the eddy-covariance (Sakai et al., 2001; Foken et al., 2006) and horizontal length scales of 1.5 times the boundary-layer height (Kaimal et al., 1976; Liu et al., 2011). Foken, (2008) pointed that eddy-covariance flux towers tend to under-sample warm convective areas, and over-sample cold areas, which can also contribute as part of the "missing" energy.

Atmospheric storage flux (Eqn. 1, term 1) does have a diurnal cycle and influences energy closure at hourly or finer temporal scale, despite of its small averaged magnitude over daily

timescale (Leuning *et al*., 2012). Especially in the case of tall towers, storage beneath the turbulent flux measurement height can comprise a substantial amount of the actual surface-atmosphere exchange at hourly resolution. Moreover, the flux footprint of turbulent flux measured at the tower top is mismatched with footprint of storage flux measured below, which is a potential source too (Schmid 1997).

For other possible sources, systematic studies on contributions to this problem have ruled out measuring errors in turbulent flux, radiation (Kohsiek et al., 2007), and ground heat (Liebethal et al., 2005; Frank *et al*., 2013; Horst *et al*., 2014; Liu *et al*., 2011), since they are not at the sufficient magnitude to close energy balance all the time at all locations. The energy consumed by photosynthesis is very small too (2-3%). While lack of full observation of horizontal advection has been shown to contribute to some of this closure (e.g., Aubinet *et al*., 2010; Barr *et al*., 2013; Nakai *et al*., 2014; Zitouna-Chebbi, 2012), it should not have a systematic direction that would lead to lack of closure at all sites.

Solutions to the energy balance closure problem are even less developed, as mesoscale eddies and storage term are hard to measure. Previous studies found that adding more towers (Steinfeld et al., 2007; Mauder et al., 2008) or expanding averaging time (Finnigan 2003; Kanda et al., 2004) can include mesoscale eddies in eddy-covariance measurements to help close energy budget (Kanda et al., 2004; Eder et al., 2015). In Chapter 3 and 4, I tested how many towers are enough to close the energy budget over regional scale heterogeneous surface, and how wavelet based flux methods and ERF upscaling can reduce observation requirement, while still adequately closing the energy budget.

1.2 Background on Eddy Covariance

Before giving the overview of the researches (Sect. 1.3) that I have done to address the abovementioned two challenges, this section gives a background of the main tool used in this dissertation: eddy-covariance technique.

Eddy-covariance flux towers are usually built 2–100 meters higher than research surface, e.g. forests, croplands, and urban, immediately above the roughness sublayer where shear stress is the dominant source of eddy transport. At the top of one flux tower mounted sonic anemometer and inferred gas analyzer (IRGA). Sonic anemometer measures potential temperature, wind speed u, v and w with respect to the Cartesian coordinates x, y, and z at 10 Hz or 20 Hz (Figure 3). IRGA measures CO_2 and H_2O concentration at the same frequency with sonic anemometers. Some towers are installed with advanced spectrometers measuring CH_4 concentration concentration continuously at high frequency. The underlying principle of eddy-covariance measurements is that when the wind flow passes by the tower, the tower is continuously measuring eddies, as if the tower is moving to measure eddies, assuming the eddies won't change too much during the time period they are passing by the tower (Taylor Hypothesis; Taylor 1938). The net surface-atmosphere exchange (NSAE) out of the control volume (Figure 4) can be derived using mass conservation and makes use of the Reynolds decomposition of relevant terms (Foken and Nappo, 2008, Stull, 1988) in the Navier-Stokes equation (Eqn. 1).

Figure 3 US-Park Falls WLEF very tall tower (left plot) in North Wisconsin with sonic anemometers (top one in the right plot) and inferred gas analyzers (IRGA, bottom one in the left plot) mounted at 30 m, 122 m, and 396 m.

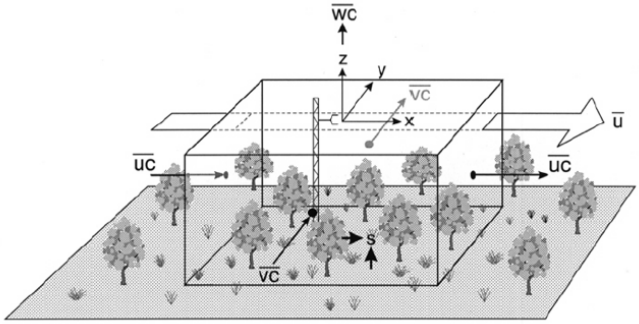


Figure 4 Cartesian control volume over a homogeneous surface patch (Finnigan et al., 2003). The coordinate axes are aligned with the mean wind vector.

$$NSAE = \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{X}}{\partial t} dz}_{I} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{u'X'}}{\partial x} dz}_{II} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{v'X'}}{\partial y} dz}_{III} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \overline{w'X'}}{\partial z} dz}_{IV} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{u}\bar{X}}{\partial x} dz}_{V} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{v}\bar{X}}{\partial y} dz}_{VI} + \underbrace{\int_0^{d_{z,m}} \frac{\partial \bar{w}\bar{X}}{\partial z} dz}_{VII}. \quad (1)$$

where X is a scalar quantity such as temperature, H_2O or CO_2 mixing ratios; u, v and w are along-, cross-, and vertical wind speeds with respect to the Cartesian coordinates x, y, and z; t is time, and $d_{z,m}$ is the measurement height. The overbars indicate a time average and primes denote turbulent fluctuations. Term I in Eqn. 1 is the rate of change of X in the vertical column below the sensor, equivalent to storage. Terms II–IV are turbulent fluxes, and terms V–VII represent advection through the layer between the surface and sensor.

Given steady state ($\partial/\partial t$=0), term I cancels from Eqn. 1. Term I is considered to be small only over the full diurnal cycle, but can be significant over shorter time intervals. Assuming horizontal homogeneity ($\partial/\partial x$=0, $\partial/\partial y$=0) implies equal horizontal inflow and outflow at opposite faces of a control volume (Fig. 2), and cancels horizontal turbulent flux (terms II–III, Eqn 1) and advection (terms V–VI, Eqn 1) from Eqn. 1. Furthermore, neglecting subsidence or convection (w≡0) cancels term VII from Eqn. 1, and hence only term IV remains.

Using dimensional analysis and similarity numbers it can be shown that the eddy covariance $w'X'$ is constant (non-dimensional) or weakly linear with height in the atmospheric surface layer to within approximately 10% (e.g., Foken, 2008). When all above conditions hold true, the total flux is then equal to the tower-measured half-hourly/hourly turbulent flux, which is the covariance of vertical wind speed and the interest scalar at half hour or an hour time scale:

$$NSAE = \overline{w'X'} \tag{2}$$

The source area or the upwind area seen by the eddy-covariance sensors is called footprint, which is a function of tower height, wind speed, wind direction, surface roughness, etc. It is worth noticing that the footprint is temporally changing with the change of shear production by changes in above ground wind speed and direction. The footprint covers different ecosystems simultaneously, e.g. wetlands and forests. Flux footprint is defined as the source area or upwind

area of the exchange/flux measured by the instruments at the top of the flux tower, which theoretically is infinite in length, but typically defined by the cumulative influence of the first 90-99% depending on the area of interest sought. The footprint is controlled by tower height, mean wind speed gradient, surface shear stress, wind direction, surface roughness, cross-wind speed, and etc. (Finnigan et al., 1996; Horst, 1979). For example, in a free-convective situation with weak winds or wind shear, , the footprint of the flux measurement is essentially a fixed area below the tower scaling with size of overturning eddies. While with a strong wind speed from east, the footprint locates at the far east side of the tower. Footprint is a critical concept in this work and for micrometeorology observations in general, as it determines the source area of flux measurements and hence help understand the surface property attributions of flux measurements.

1.3 Overview of this dissertation

This dissertation touches on three key questions:

1) Is it possible to rectify location bias, by extending the application of the environmental response function (ERF) upscaling method, from the airborne eddy-covariance to tower eddy-covariance measurements?
2) Can ERF close energy budget in eddy-covariance observations by resolving the not-easy-measured storage flux and vertical advection flux?
3) How does ERF perform compared to previous scaling approaches using one or multiple towers?

Chapter 2–4 target on these three questions separately. Chapter 2 develops the environmental response functions (ERFs) upscaling method for eddy-covariance tower measurements to project turbulent flux maps at regional scale and hourly temporal resolution. Further, it assesses

the uncertainty budget of ERF, and evaluates decomposed fluxes against other independent observations and estimates of regional flux from pervious upscaling methods.

Chapter 3 further develops tower-ERF approach to expand its application to not only turbulent, but also vertical advection and storage flux, for the first time practically implements ERF to retrieve "virtual control volume (VCV)" net surface-atmosphere exchange, tests the assumptions made in theoretical ERF-VCV, and explores the hypothesis that ERF-VCV-rectified net surface-atmosphere exchange can resolve "storage flux" and "vertical advection" issues.

Chapter 4 tests and validates ERF approach in an LES experiment with known heterogeneous surface forcing, compares the performance of ERF with a traditional EC approach of adding multiple towers in perspectives of rectifying location bias, closing energy budget, and spatial pattern retrieval.

Major findings from Chapters 2–4 and directions for future research are summarized in Chapter 5.

1.4 Reference

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., . . . Tans, P. P. (2014). CO2, CO, and CH4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques, 7*(2), 647-687. doi:10.5194/amt-7-647-2014

Antonarakis, A., Munger, J., & Moorcroft, P. (2014). Imaging spectroscopy-and lidar-derived estimates of canopy composition and structure to improve predictions of forest carbon fluxes and ecosystem dynamics. *Geophysical Research Letters, 41*(7), 2535-2542. doi:10.1002/2013GL058373

Aubinet, M., Feigenwinter, C., Heinesch, B., Bernhofer, C., Canepa, E., Lindroth, A., . . . Van Gorsel, E. (2010). Direct advection measurements do not help to solve the night-time CO2 closure problem: Evidence from three different forests. *Agricultural and Forest Meteorology, 150*(5), 655-664. doi:10.1016/j.agrformet.2010.01.016

Avissar, R., & Liu, Y. (1996). Three-dimensional numerical study of shallow convective clouds and precipitation induced by land surface forcing. *Journal of Geophysical Research: Atmospheres, 101*(D3), 7499-7518.

Bakwin, P. S., Tans, P. P., Hurst, D. F., & Zhao, C. L. (1998). Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program. *Tellus Series B-Chemical and Physical Meteorology, 50*(5), 401-415. doi:10.1034/j.1600-0889.1998.t01-4-00001.x

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. *Bulletin of the American Meteorological Society, 82*(11), 2415-2434. doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Barr, A. G., Richardson, A. D., Hollinger, D. Y., Papale, D., Arain, M. A., Black, T. A., . . . Schaeffer, K. (2013). Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies. *Agricultural and Forest Meteorology, 171-172*, 31-45. doi:10.1016/j.agrformet.2012.11.023

Betts, A. K., & Viterbo, P. (2005). Land-surface, boundary layer, and cloud-field coupling over the southwestern Amazon in ERA-40. *Journal of Geophysical Research-Atmospheres, 110*(D14), 15. doi:10.1029/2004jd005702

Bohrer, G., Katul, G. G., Walko, R. L., & Avissar, R. (2009). Exploring the Effects of Microscale Structural Heterogeneity of Forest Canopies Using Large-Eddy Simulations. *Boundary-Layer Meteorology, 132*(3), 351-382. doi:10.1007/s10546-009-9404-4

Bonan, G. B. (2008). *Ecological climatology: concepts and applications*. Cambridge, UK: Cambridge University Press

Caughey, S., Wyngaard, J., & Kaimal, J. (1979). Turbulence in the evolving stable boundary layer. *Journal of the Atmospheric Sciences, 36*(6), 1041-1052.

Chen, F., & Avissar, R. (1994). The impact of land-surface wetness heterogeneity on mesoscale heat fluxes. *Journal of Applied Meteorology, 33*(11), 1323-1340.

Davis, K. J., Bakwin, P. S., Yi, C. X., Berger, B. W., Zhao, C. L., Teclaw, R. M., & Isebrands, J. G. (2003). The annual cycles of CO2 and H2O exchange over a northern mixed forest as observed from a very tall tower. *Global Change Biology, 9*(9), 1278-1293. doi:10.1046/j.1365-2486.2003.00672.x

Desai, A. R., Davis, K. J., Senff, C. J., Ismail, S., Browell, E. V., Stauffer, D. R., & Reen, B. P. (2006). A case study on the effects of heterogeneous soil moisture on mesoscale boundary-layer structure in the Southern Great Plains, USA Part I: Simple prognostic model. *Boundary-Layer Meteorology, 119*(2), 195-238. doi:10.1007/s10546-005-9024-6

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., . . . Stauch, V. J. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and Forest Meteorology, 148*(6-7), 821-838. doi:10.1016/j.agrformet.2007.11.012

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. *Agricultural and Forest Meteorology, 201*, 61-75. doi:10.1016/j.agrformet.2014.10.017

Drobinski, P., Brown, R. A., Flamant, P. H., & Pelon, J. (1998). Evidence of organized large eddies by ground-based Doppler lidar, sonic anemometer and sodar. *Boundary-Layer Meteorology, 88*(3), 343-361. doi:10.1023/a:1001167212584

Eder, F., De Roo, F., Rotenberg, E., Yakir, D., Schmid, H. P., & Mauder, M. (2015). Secondary circulations at a solitary forest surrounded by semi-arid shrubland and their impact on eddy-covariance measurements. *Agricultural and Forest Meteorology, 211–212*, 115-127. doi:http://dx.doi.org/10.1016/j.agrformet.2015.06.001

Eder, F., Roo, F., Kohnert, K., Desjardins, R. L., Schmid, H. P., & Mauder, M. (2014). Evaluation of Two Energy Balance Closure Parametrizations. *Boundary-Layer Meteorology, 151*(2), 195-219. doi:10.1007/s10546-013-9904-0

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x

Ershadi, A., McCabe, M., Evans, J., & Walker, J. (2013). Effects of spatial aggregation on the multi-scale estimation of evapotranspiration. *Remote Sensing of Environment, 131*, 51-62. doi:10.1016/j.rse.2012.12.007

Finnigan, J. J., Clement, R., Malhi, Y., Leuning, R., & Cleugh, H. A. (2003). A re-evaluation of long-term flux measurement techniques - Part I: Averaging and coordinate rotation. *Boundary-Layer Meteorology, 107*(1), 1-48. doi:10.1023/a:1021554900225

Foken, T. (2008). The energy balance closure problem: an overview. *Ecological Applications, 18*(6), 1351-1367. doi:10.1890/06-0922.1

Foken, T., Aubinet, M., Finnigan, J. J., Leclerc, M. Y., Mauder, M., & Paw U, K. T. (2011). Results of a panel discussion about the energy balance closure correction for trace gases. *Bulletin of the American Meteorological Society, 92*(4), ES13-ES18. doi:10.1175/2011BAMS3130.1

Foken, T., & Nappo, C. J. (2008). *Micrometeorology*: Springer

Foken, T., Wimmer, F., Mauder, M., Thomas, C., & Liebethal, C. (2006). Some aspects of the energy balance closure problem. *Atmospheric Chemistry and Physics, 6*, 4395-4402.

Frank, J. M., Massman, W. J., & Ewers, B. E. (2013). Underestimates of sensible heat flux due to vertical velocity measurement errors in non-orthogonal sonic anemometers. *Agricultural and Forest Meteorology, 171*, 72-81. doi:10.1016/j.agrformet.2012.11.005

Grimmond, C. S. B., Blackett, M., Best, M. J., Baik, J. J., Belcher, S. E., Beringer, J., . . . Zhang, N. (2011). Initial results from Phase 2 of the international urban energy balance model comparison. *International Journal of Climatology, 31*(2), 244-272. doi:10.1002/joc.2227

Hilker, T., Hall, F. G., Coops, N. C., Collatz, J. G., Black, T. A., Tucker, C. J., . . . Grant, N. (2013). Remote sensing of transpiration and heat fluxes using multi-angle observations. *Remote Sensing of Environment, 137*, 31-42. doi:10.1016/j.rse.2013.05.023

Hill, T., Quaife, T., & Williams, M. (2011). A data assimilation method for using low-resolution Earth observation data in heterogeneous ecosystems. *Journal of Geophysical Research: Atmospheres (1984–2012), 116*(D8).

Hill, T. C., Williams, M., & Moncrieff, J. B. (2008). Modeling feedbacks between a boreal forest and the planetary boundary layer. *Journal of Geophysical Research-Atmospheres, 113*(D15). doi:10.1029/2007jd009412

Horst, T. W., Semmer, S. R., & Maclean, G. (2015). Correction of a Non-orthogonal, Three-Component Sonic Anemometer for Flow Distortion by Transducer Shadowing. *Boundary-Layer Meteorology, 155*(3), 371-395. doi:10.1007/s10546-015-0010-3

Hutjes, R., Vellinga, O., Gioli, B., & Miglietta, F. (2010). Dis-aggregation of airborne flux measurements using footprint analysis. *Agricultural and Forest Meteorology, 150*(7), 966-983. doi:10.1016/j.agrformet.2010.03.004

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., . . . De Jeu, R. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature, 467*(7318), 951-954. doi:10.1038/nature09396

Kaimal, J., Wyngaard, J., Haugen, D., Coté, O., Izumi, Y., Caughey, S., & Readings, C. (1976). Turbulence structure in the convective boundary layer. *Journal of the Atmospheric Sciences, 33*(11), 2152-2169.

Kampe, T. U., McCorkel, J., Hamlin, L., Green, R. O., Krause, K. S., & Johnson, B. R. (2011). Progress in the development of airborne remote sensing instrumentation for the National Ecological Observatory Network. In W. Gao, T. J. Jackson, J. Wang, & N. B. Chang (Eds.), *Remote Sensing and Modeling of Ecosystems for Sustainability Viii* (Vol. 8156).

Kanda, M., Inagaki, A., Letzel, M. O., Raasch, S., & Watanabe, T. (2004). LES study of the energy imbalance problem with Eddy covariance fluxes. *Boundary-Layer Meteorology, 110*(3), 381-404. doi:10.1023/B:BOUN.0000007225.45548.7a

Kang, S. L. (2009). Temporal Oscillations in the Convective Boundary Layer Forced by Mesoscale Surface Heat-Flux Variations. *Boundary-Layer Meteorology, 132*(1), 59-81. doi:10.1007/s10546-009-9391-5

Kang, S. L., Davis, K. J., & LeMone, M. (2007). Observations of the ABL structures over a heterogeneous land surface during IHOP_2002. *Journal of Hydrometeorology, 8*(2), 221-244. doi:10.1175/jhm567.1

Kang, S. L., & Lenschow, D. H. (2014). Temporal Evolution of Low-Level Winds Induced by Two-dimensional Mesoscale Surface Heat-Flux Heterogeneity. *Boundary-Layer Meteorology, 151*(3), 501-529. doi:10.1007/s10546-014-9912-8

Kobayashi, T., & Nishizawa, N. K. (2012). Iron uptake, translocation, and regulation in higher plants. *Annual review of plant biology, 63*, 131-152. doi:10.1146/annurev-arplant-042811-105522

Kohsiek, E., Liebethal, C., Foken, T., Vogt, R., Oncley, P., Bernhofer, C., & Debruin, R. (2007). The energy balance experiment EBEX-2000. Part III: Behaviour and quality of the radiation measurements. *Boundary-Layer Meteorology, 123*(1), 55-75. doi:10.1007/sl0546-006-9l35-8

Lenschow, D., & Stephens, P. (1980). The role of thermals in the convective boundary layer. *Boundary-Layer Meteorology, 19*(4), 509-532.

Leuning, R., van Gorsel, E., Massman, W. J., & Isaac, P. R. (2012). Reflections on the surface energy imbalance problem. *Agricultural and Forest Meteorology, 156*, 65-74. doi:10.1016/j.agrformet.2011.12.002

Liebethal, C., Huwe, B., & Foken, T. (2005). Sensitivity analysis for two ground heat flux calculation approaches. *Agricultural and Forest Meteorology, 132*(3-4), 253-262. doi:10.1016/j.agrformet.2005.08.001

Liu, S. M., Xu, Z. W., Wang, W. Z., Jia, Z. Z., Zhu, M. J., Bai, J., & Wang, J. M. (2011). A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. *Hydrology and Earth System Sciences, 15*(4), 1291-1306. doi:10.5194/hess-15-1291-2011

Lund, T. S., Wu, X., & Squires, K. D. (1998). Generation of turbulent inflow data for spatially-developing boundary layer simulations. *Journal of Computational Physics, 140*(2), 233-258.

Mahrt, L. (1998). Flux sampling errors for aircraft and towers. *Journal of Atmospheric and Oceanic Technology, 15*(2), 416-429. doi:10.1175/1520-0426(1998)015<0416:fsefaa>2.0.co;2

Manoli, G., Domec, J. C., Novick, K., Oishi, A. C., Noormets, A., Marani, M., & Katul, G. (2016). Soil-Plant-Atmosphere Conditions Regulating Convective Cloud Formation Above Southeastern US Pine Plantations. *Global Change Biology*.

Maronga, B., & Raasch, S. (2013). Large-Eddy Simulations of Surface Heterogeneity Effects on the Convective Boundary Layer During the LITFASS-2003 Experiment. *Boundary-Layer Meteorology, 146*(1), 17-44. doi:10.1007/s10546-012-9748-z

Mauder, M., Desjardins, R. L., & MacPherson, I. (2007). Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem. *Journal of Geophysical Research: Atmospheres (1984–2012), 112*(D13).

Mauder, M., Eder, F., De Roo, F., Brugger, P., Schmid, H. P., Rotenberg, E., & Yakir, D. (2015). *Large-eddy transport in the surface layer over heterogeneous terrain*. Paper presented at the 48th AGU annual Fall Meeting, San Francisco, U.S.A.

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. doi:10.5194/bg-10-2193-2013

Metzger, S., Xu, K., Desai, A. R., Taylor, J. R., Kljun, N., Fox, A., & Kampe, T. (2014). *Mapping the flux field around eddy-covariance measurements*. Paper presented at the 31st AMS Conference on Agricultural and Forest Meteorology, Portland, U.S.A.

Metzger, S.: Surface-atmosphere exchange in a box: Making it a suitable representation for in-situ observations, 32nd Conference on Agricultural and Forest Meteorology, Salt Lake City, U.S.A., 20 - 24 June, 2016.

Metzger, S. (2017). Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations. *Agricultural and Forest Meteorology*(Massman special issue).

Moeng, C.-H. (1984). A large-eddy-simulation model for the study of planetary boundary-layer turbulence. *Journal of the Atmospheric Sciences, 41*(13), 2052-2062.

Musinsky, J., Wasser, L., Kampe, T., Leisso, N., Krause, K., Petroy, S., . . . Serbin, S. (2013). *Developing a Scalable Remote Sampling Design for the NEON Airborne Observation Platform (AOP).* Paper presented at the AGU Fall Meeting Abstracts.

Nakai, T., Iwata, H., Harazono, Y., & Ueyama, M. (2014). An inter-comparison between Gill and Campbell sonic anemometers. *Agricultural and Forest Meteorology, 195*, 123-131. doi:10.1016/j.agrformet.2014.05.005

Neale, C. M. U., Geli, H. M. E., Kustas, W. P., Alfieri, J. G., Gowda, P. H., Evett, S. R., . . . Howell, T. A. (2012). Soil water content estimation using a remote sensing based hybrid evapotranspiration modeling approach. *Advances in Water Resources, 50*, 152-161. doi:10.1016/j.advwatres.2012.10.008

Pleim, J. E., & Xiu, A. (1995). Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. *Journal of Applied Meteorology, 34*(1), 16-32.

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. *Advances in Water Resources, 50*, 106-119. doi:10.1016/j.advwatres.2012.07.014

Raasch, S., Schroter, M., Ams, & Ams. (2002). *A large-eddy simulation model performing on massively parallel computers*. Boston: Amer Meteorological Society

Rihani, J. F., Chow, F. K., & Maxwell, R. M. (2015). Isolating effects of terrain and soil moisture heterogeneity on the atmospheric boundary layer: Idealized simulations to diagnose land-atmosphere feedbacks. *Journal of Advances in Modeling Earth Systems, 7*(2), 915-937. doi:10.1002/2014ms000371

Sakai, R. K., Fitzjarrald, D. R., & Moore, K. E. (2001). Importance of low-frequency contributions to eddy fluxes observed over rough surfaces. *Journal of Applied Meteorology, 40*(12), 2178-2192. doi:10.1175/1520-0450(2001)040<2178:iolfct>2.0.co;2

Schlegel, F., Stiller, J., Bienert, A., Maas, H.-G., Queck, R., & Bernhofer, C. (2014). Large-Eddy Simulation Study of the Effects on Flow of a Heterogeneous Forest at Sub-Tree Resolution. *Boundary-Layer Meteorology, 154*(1), 27-56. doi:10.1007/s10546-014-9962-y

Schmid, H. (1994). *Probabilistic part-of-speech tagging using decision trees.* Paper presented at the Proceedings of the international conference on new methods in language processing.

Schmid, H. P. (1997). Experimental design for flux measurements: matching scales of observations and fluxes. *Agricultural and Forest Meteorology, 87*(2-3), 179-200. doi:10.1016/s0168-1923(97)00011-7

Schmid, H. P. (2002). Footprint modeling for vegetation atmosphere exchange studies: a review and perspective. *Agricultural and Forest Meteorology, 113*(1), 159-183.

Shrestha, P., Sulis, M., Simmer, C., & Kollet, S. (2015). Impacts of grid resolution on surface energy fluxes simulated with an integrated surface-groundwater flow model. *Hydrology and Earth System Sciences, 19*(10), 4317-4326.

Steinfeld, G., Letzel, M. O., Raasch, S., Kanda, M., & Inagaki, A. (2007). Spatial representativeness of single tower measurements and the imbalance problem with eddy-covariance fluxes: results of a large-eddy simulation study. *Boundary-Layer Meteorology, 123*(1), 77-98.

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. *Agricultural and Forest Meteorology, 171*, 137-152. doi:10.1016/j.agrformet.2012.11.004

Stull, R. B. (1988). *An introduction to boundary layer meteorology* (Vol. 13): Springer Science & Business Media

Sulis, M., Langensiepen, M., Shrestha, P., Schickling, A., Simmer, C., & Kollet, S. J. (2015). Evaluating the influence of plant-specific physiological parameterizations on the partitioning of land surface energy fluxes. *Journal of Hydrometeorology, 16*(2), 517-533.

Sun, J. L., Lenschow, D. H., Mahrt, L., Crawford, T. L., Davis, K. J., Oncley, S. P., . . . Desjardins, R. L. (1997). Lake-induced atmospheric circulations during BOREAS. *Journal of Geophysical Research-Atmospheres, 102*(D24), 29155-29166. doi:10.1029/97jd01114

Traeumner, K., Damian, T., Stawiarski, C., & Wieser, A. (2015). Turbulent Structures and Coherence in the Atmospheric Surface Layer. *Boundary-Layer Meteorology, 154*(1), 1-25. doi:10.1007/s10546-014-9967-6

Troldborg, N., Sørensen, J. N., Mikkelsen, R., & Sørensen, N. N. (2014). A simple atmospheric boundary layer model applied to large eddy simulations of wind turbine wakes. *Wind Energy, 17*(4), 657-669.

van Heerwaarden, C. C., Mellado, J. P., & De Lozar, A. (2014). Scaling Laws for the Heterogeneously Heated Free Convective Boundary Layer. *Journal of the Atmospheric Sciences, 71*(11), 3975-4000. doi:10.1175/jas-d-13-0383.1

Vilà-Guerau de Arellano, J., Jonker, H., Pino, D., Ten Brink, H. M., Chaumerliac, N., Faloona, I., . . . Stutz, J. (2007). The Role of Atmospheric Boundary Layer Processes in Atmospheric Chemistry. *Bulletin of the American Meteorological Society, 88*(8), 1245-1248. doi:10.1175/bams-88-8-1245

Wang, K., & Dickinson, R. E. (2012). A REVIEW OF GLOBAL TERRESTRIAL EVAPOTRANSPIRATION: OBSERVATION, MODELING, CLIMATOLOGY, AND CLIMATIC VARIABILITY. *Reviews of Geophysics, 50*. doi:10.1029/2011rg000373

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. *Journal of Geophysical Research: Biogeosciences, 111*(G2). doi:10.1029/2005JG000093

Wilczak, J. M., Oncley, S. P., & Stage, S. A. (2001). Sonic anemometer tilt correction algorithms. *Boundary-Layer Meteorology, 99*(1), 127-150. doi:10.1023/A:1018966204465

Wilczak, J. M., & Tillman, J. E. (1980). THE 3-DIMENSIONAL STRUCTURE OF CONVECTION IN THE ATMOSPHERIC SURFACE-LAYER. *Journal of the Atmospheric Sciences, 37*(11), 2424-2443. doi:10.1175/1520-0469(1980)037<2424:ttdsoc>2.0.co;2

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. *Journal of Geophysical Research, 116*(G3). doi:10.1029/2010jg001568

Xiao, J., Ollinger, S. V., Frolking, S., Hurtt, G. C., Hollinger, D. Y., Davis, K. J., . . . Suyker, A. E. (2014). Data-driven diagnostics of terrestrial carbon dynamics over North America. *Agricultural and Forest Meteorology, 197*, 142-157. doi:10.1016/j.agrformet.2014.06.013

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., . . . Torn, M. S. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology, 148*(11), 1827-1847. doi:10.1016/j.agrformet.2008.06.015

Xu, K., Metzger, S., & Desai, A. R. (2015). *Mapping surface-atmosphere exchange by using environmental response function for both turbulent and storage eddy-covariance fluxes*. Paper presented at the 48th AGU annual Fall Meeting, San Francisco, U.S.A.

Xu, K., Metzger, S., & Desai, A. R. (2017). Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions. *Agricultural and Forest Meteorology*.

Xu, K., Metzger, S., Kljun, N., Taylor, J. R., and Desai, A. R.: How to consistently inform NEON's land surface model with tower-based eddy covariance flux observations? A novel approach to spatio-temporal rectification, 99th ESA Annual Meeting, Sacramento, U.S.A., 10 - 15 August, 2014.

Zitouna-Chebbi, R., Prevot, L., Jacob, F., Mougou, R., & Voltz, M. (2012). Assessing the consistency of eddy covariance measurements under conditions of sloping topography within a hilly agricultural catchment. *Agricultural and Forest Meteorology, 164*, 123-135. doi:10.1016/j.agrformet.2012.05.010

2 Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions[1]

Keywords: Upscale; eddy-covariance; regional flux map; land-atmosphere interaction; footprint bias; fine spatio-temporal resolution

Highlights:

- Environmental Response Function (ERF) maps surface-atmosphere exchanges.
- ERF is capable of upscaling tower eddy covariance observations to regional scale.
- The regional flux maps de-bias the footprint variation in tower observations.
- ERF provides unprecedented spatio-temporal attribution and resolution.

2.1 Introduction

Earth system models (ESMs) have been developed and are widely used to understand impacts of global climate change (Le Quéré et al., 2012, 2013, 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/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

[1]This chapter is modified and excerpted with permission from the published from Xu, K., Metzger, S., & Desai, A. R. (2017). Upscaling tower-observed turbulent exchange at fine spatio-temporalresolution using environmental response functions. *Agricultural and Forest Meteorology, 232*, 10-22. doi:10.1016/j.agrformet.2016.07.019.

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, 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).

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 (Jacinth 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, 2008; 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, 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 (Sect. 2.2.1). The methodology of ERF scaling approach and associated uncertainty algorithm are described in Sect. 2.2.2 and Sect. 2.2.3. We present the extracted ERFs, scaled flux grids and the associated uncertainty budget in Sect. 2.3.1, 2.3.2 and 2.3.3, separately. A comparison of ERF scaling approach and its outcomes with in-situ measurements and other scaling approaches is performed in Sect. 2.3.4. In Sect. 5.2, 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 Sect. 2.3.6, and conclusions in Sect. 2.4.

2.2 Materials and methods

2.2.1 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. 5). 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 (Figure 6a), 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 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 (Figure 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 hour 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.

2.2.2 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 (Sects. 2.2.2.2–2.2.2.3, Fig. 8); ii) extracting ERFs between flux responses and drivers using machine learning (Sect. 2.2.2.4, Fig. 9, Fig. 10); iii) projecting turbulent exchange grids over the target area based

on extracted ERFs and key drivers in each grid cell (Sect. 2.2.2.4, 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 (Sect. 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 (Sect. 2.2.2.2). The corresponding environmental drivers included two types: temporally explicit meteorological variables and spatially resolved land surface properties (Sect. 2.2.3). Lastly, ERFs was extracted and used for projection in Sect.2.2.2.4.

2.2.2.1 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. 7). 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 ± 88 m, e.g. Strunin et al., 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.

2.2.2.2 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 et al., 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 Sect. 2.2.2.1, 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, 2008b, 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, $u^* > 0.2\ m\ s^{-1}$ was used to select H at 30 m and 122 m levels, $H > -20$ W m^{-2} at 30 m was used to select fluxes at 122 m, and $w^* > 0.8\ m\ s^{-1}$ 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.

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

After preprocessing steps mentioned in Sect. 2.2.2.2, 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 artificial 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 (Sect. 2.2.3, 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 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.

2.2.2.4 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 $mmol\ mol^{-1}$) 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 20×20 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 (Sect. 2.2.3, 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.

2.2.3 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 Sect. 2.2.2.2. 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 ERF 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.

2.3 Results and discussion

In Sect. 2.3.1, 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 Sect. 2.3.2, the projection results from ERFs and summarized probability density functions (PDFs) are shown and discussed. Sects. 2.3.3–2.3.6 present and discuss the uncertainty, comparison with other upscaling approaches, as well as prospect and limitation of ERF scaling, respectively.

2.3.1 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. 7, the footprint changed from southwest narrow strip (Fig. 7a) over small closed area (Fig. 7b) to south larger area (Fig. 7c) 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%±4%, -8.7%±4% and +7.5%±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 Figure 8, 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 Figure 8). 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. 9 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. 9 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 (Sect. 2.2.2.4). 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. 9 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 that θ acts as a proxy for the actual radiative forcing due to the presence of clouds (ref.), 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. 10a, 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. 10a). 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. 10c). 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.

2.3.2 Spatial-temporally resolved flux maps

Figure 11 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 1% (average footprint area relative to 20 km x 20 km) to 56%±13%, 51% ± 6%, and 51% ± 13% for 20×20 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. 12) 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 Wm^{-2}, 44.8 Wm^{-2} and –

94.4 gC/month/m^2, compared to the flux observations, 39.6 Wm^{-2}, 49.2 Wm^{-2}, and –80.4 gC/month/m^2. One standard deviation of spatio-temporal variation of projected *H*, LE, and CO_2 flux were 101 Wm^{-2}, 73 Wm^{-2}, and 322 gC/month/m^2, smaller than observations, 124 Wm^{-2}, 92 Wm^{-2}, 332 gC/month/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 (Figure 6). 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 291K-295K (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, 2008a). 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.

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

The latter 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.

2.3.4 Comparison with measurements and other studies

The averaged ERF-projection over 3 km x 3 km region centered by WCr explained 76%, 80% of the variation of H observed at WCr when using WCr and WLEF meteorological drivers for projection, respectively (Figure 13). ERF-projection underestimated (slope) observations by 17% and 9% with residual standard deviation at 7.49 Wm^{-2} and 0.99 Wm^{-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 Figure 13b 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±31 gC m^{-2} $month^{-1}$ is consistent with a footprint-debiased estimate -86 ± 18 gC m^{-2} $month^{-1}$ and a multi-tower synthesis aggregation -97±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 ± 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 3×3 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.

2.3.5 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 Figure 11, 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 (Figure 11a, 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 (Figure 11b). Therefore, ERF-rectified flux grids mitigate the pitfall of limited and varying spatial representativeness when using observations to inform a model during execution.

2.3.6 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.

2.4 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 rectifies 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.

2.5 Acknowledgement

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.

2.6 References

Ammann, C., Brunner, A., Spirig, C., & Neftel, A. (2006). Technical note: Water vapour concentration and flux measurements with PTR-MS. *Atmospheric Chemistry and Physics, 6*(12), 4643-4651. doi: 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. *Journal of Climate, 26*(15), 5289-5314. doi: 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. *Bulletin of the American Meteorological Society, 82*(11), 2415-2434. doi: 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. *Agricultural and Forest Meteorology, 151*(3), 394-405. doi: 10.1016/j.agrformet.2010.12.001

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., . . . Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research: Biogeosciences (2005–2012), 116*(G2). doi: 10.1029/2010jg001593

Brock, F. V. (1986). A nonlinear filter to remove impulse noise from meteorological data. *Journal of Atmospheric and Oceanic Technology, 3*(1), 51-58. doi: 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. doi: 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. *Journal of the Atmospheric Sciences, 28*(2), 181-189. doi: 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. *Geoscientific Model Development, 4*(4), 1051-1075. doi:10.5194/gmd-4-1051-2011

Croux, C., & Rousseeuw, P. J. (1992). *Time-efficient algorithms for two highly robust estimators of scale*: Springer.

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

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. *Journal of Geophysical Research: Biogeosciences (2005–2012), 115*(G2). doi: 10.1029/2009JG001122

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. *Agricultural and Forest Meteorology, 148*(2), 288-308. doi: 10.1016/j.agrformet.2007.08.001

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. *Agricultural and Forest Meteorology, 201*, 61-75. doi: 10.1016/j.agrformet.2014.10.017

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. *Journal of Geophysical Research: Biogeosciences, 119*(3), 286-300. doi: 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. *Climate Dynamics, 40*(9-10), 2123-2165. doi:10.1007/s00382-012-1636-1

Foken, T. (2008). The energy balance closure problem: an overview. *Ecological Applications, 18*(6), 1351-1367. doi: 10.1890/06-0922.1

Foken, T., & Nappo, C. J. (2008). *Micrometeorology*: Springer.

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. *Agricultural and Forest Meteorology, 140*(1), 152-170. doi: 10.1016/j.agrformet.2006.01.012

Högström, U. (1988). Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation *Topics in Micrometeorology. A Festschrift for Arch Dyer* (pp. 55-78): Springer.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., . . . Wickham, J. (2007). Completion of the 2001 national land cover database for the counterminous United States. *Photogrammetric Engineering and Remote Sensing, 73*(4), 337. doi: 10.1007/s00267-010-9536-9

Hutjes, R., Vellinga, O., Gioli, B., & Miglietta, F. (2010). Dis-aggregation of airborne flux measurements using footprint analysis. *Agricultural and Forest Meteorology, 150*(7), 966-983. doi: 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. doi: 10.1016/j.geoderma.2006.04.020

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., . . . De Jeu, R. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature, 467*(7318), 951-954. doi: 10.1038/nature09396

Kljun, N., Calanca, P., Rotach, M., & Schmid, H. (2004). A simple parameterisation for flux footprint predictions. *Boundary-Layer Meteorology, 112*(3), 503-523. doi:10.1023/B:BOUN.0000030653.71031.96

Le Quéré, C., Moriarty, R., Andrew, R., Peters, G., Ciais, P., Friedlingstein, P., . . . Arneth, A. (2014). Global carbon budget 2014. *Earth System Science Data Discussions, 7*(2), 521-610. doi: 10.5194/essdd-7-521-2014

Le Quéré, C., Peters, G., Andres, R., Andrew, R., Boden, T., Ciais, P., . . . Moriarty, R. (2013). Global carbon budget 2013. *Earth Syst. Sci. Data Discuss, 6*(2), 689-760. doi: 10.5194/essdd-6-689-2013

Lenschow, D., Mann, J., & Kristensen, L. (1994). How long is long enough when measuring fluxes and other turbulence statistics? *Journal of Atmospheric and Oceanic Technology, 11*(3), 661-673. doi: 10.1175/1520-0426(1994)011<0661:hlilew>2.0.co;2

Lenschow, D. H., & Stankov, B. B. (1986). Length scales in the convective boundary layer. *Journal of the Atmospheric Sciences, 43*(12), 1198-1209. doi: 10.1007/s10546-004-2165-1

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. doi: 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. doi: 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 Biology, 9*(4), 525-535. doi: 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. *Advances in Water Resources, 50*, 106-119. doi: 10.1016/j.advwatres.2012.07.014

Quéré, C. L., Andres, R. J., Boden, T., Conway, T., Houghton, R., House, J. I., . . . Ahlström, A. (2013). The global carbon budget 1959–2011. *Earth System Science Data, 5*(1), 165-185. doi: 10.5194/essd-5-165-2013

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 Biology, 18*(2), 566-584. doi: 10.1111/j.1365-2486.2011.02562.x

Rousseeuw, P. J., & Verboven, S. (2002). Robust estimation in very small samples. *Computational Statistics & Data Analysis, 40*(4), 741-758. doi: 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. *Journal of Geophysical Research: Biogeosciences, 117*(G3). doi: 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. *Boundary-Layer Meteorology, 26*(1), 81-93. doi: 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. *Journal of Geophysical Research: Biogeosciences, 115*(G3). doi: 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. *Journal of Geophysical Research: Atmospheres, 120*(4), 1348-1360. doi: 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. *Agricultural and Forest Meteorology, 171*, 137-152. doi: 10.1016/j.agrformet.2012.11.004

Sulkava, M., Luyssaert, S., Zaehle, S., & Papale, D. (2011). Assessing and improving the representativeness of monitoring networks: The European flux tower network example. *Journal of Geophysical Research: Biogeosciences, 116*(G3). doi: 10.1029/2010jg001562

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society, 79*(1), 61-78. doi: 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. *Journal of Geophysical Research: Biogeosciences, 111*(G2). doi: 10.1029/2005JG000093

Wilczak, J. M., Oncley, S. P., & Stage, S. A. (2001). Sonic anemometer tilt correction algorithms. *Boundary-Layer Meteorology, 99*(1), 127-150. doi: 10.1023/A: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. doi: 10.5194/bg-6-1341-2009

Xiao, J., Chen, J., Davis, K. J., & Reichstein, M. (2012). Advances in upscaling of eddy covariance measurements of carbon and water fluxes. *Journal of Geophysical Research: Biogeosciences, 117*(G1). doi: 10.1029/2011jg001889

Xiao, J., Davis, K. J., Urban, N. M., & Keller, K. (2014). Uncertainty in model parameters and regional carbon fluxes: A model-data fusion approach. *Agricultural and Forest Meteorology, 189*, 175-186. doi: 10.1016/j.agrformet.2014.01.022

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. *Journal of Geophysical Research, 116*(G3). doi: 10.1029/2010jg001568

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. *Agricultural and Forest Meteorology, 197*, 142-157. doi: 10.1016/j.agrformet.2014.06.013

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*: CRC Press.

2.7 Tables

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%)**

Table 2. ERF-projected CO_2 net ecosystem exchange [gC m^{-2} $month^{-1}$] for major land covers in the 20 km x 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–−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	-

2.8 Figures

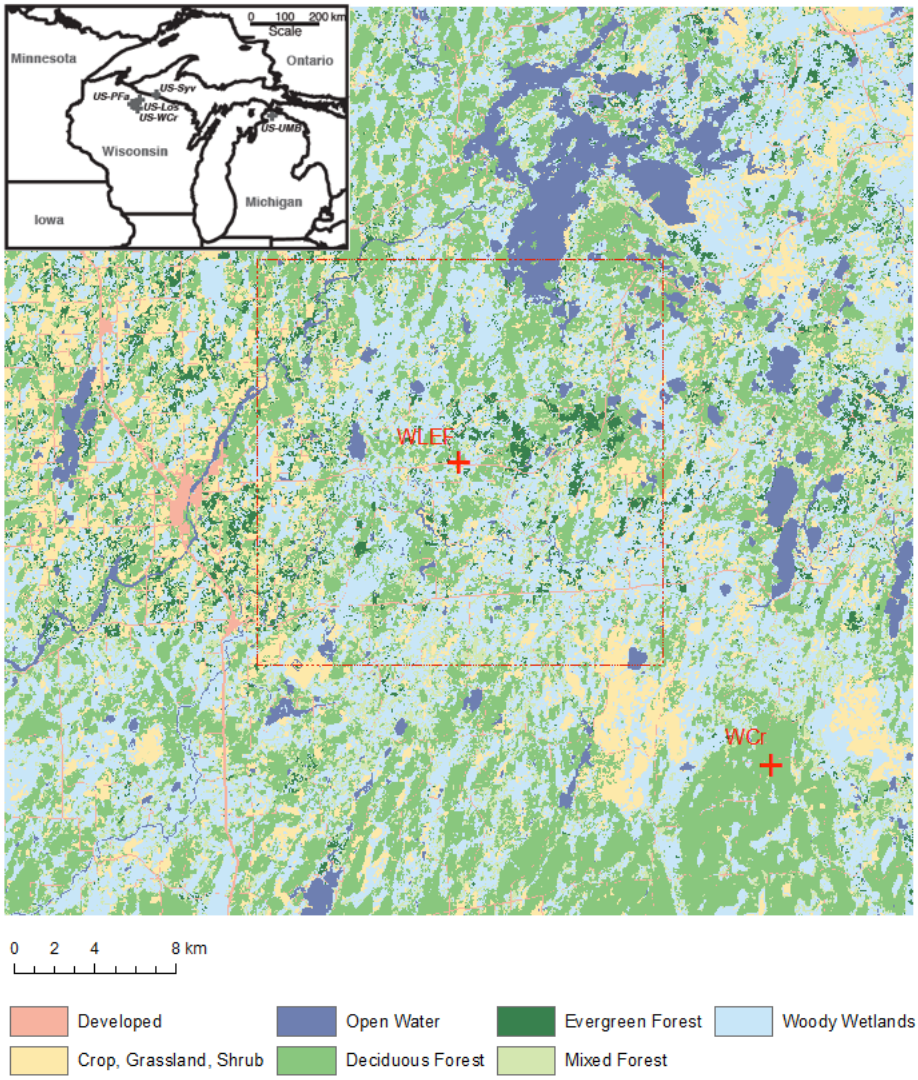
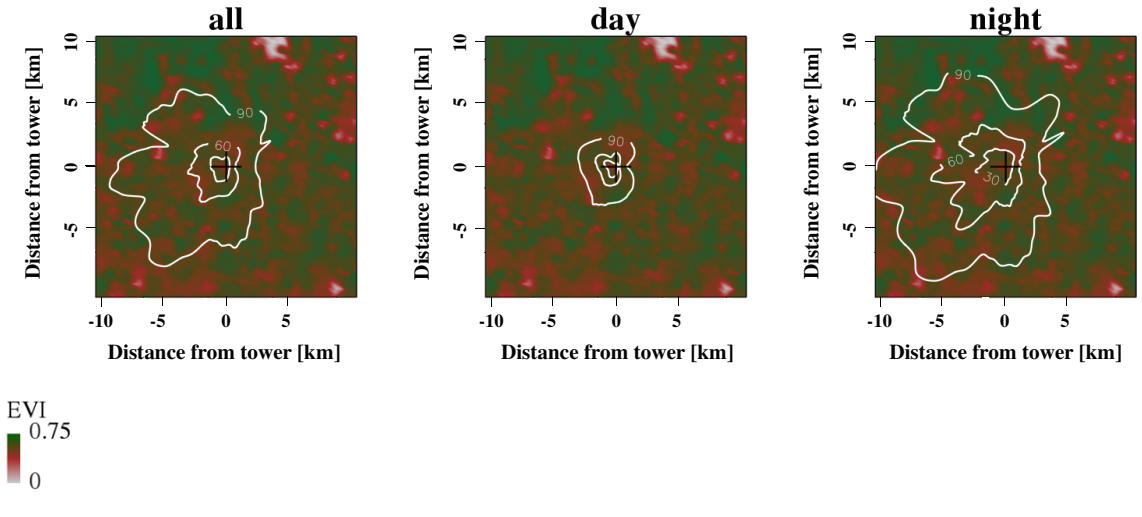


Figure 5 Land cover map (National Land Cover Database 2011) of Park Fall tower and Willow Creek tower sites (+) and 20×20 km^2 target domain (within red dashed line). The land cover data were obtained from National Land Cover Database 2011 (NLCD 2011) at spatial resolution of 30 m. NLCD 2011 is based 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.



a b c

Figure 6 Footprint climatology (30%, 60% and 90%, white contour lines) for 122 m level measurements superimposed over MODIS enhanced vegetation index (EVI) within 20×20 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).

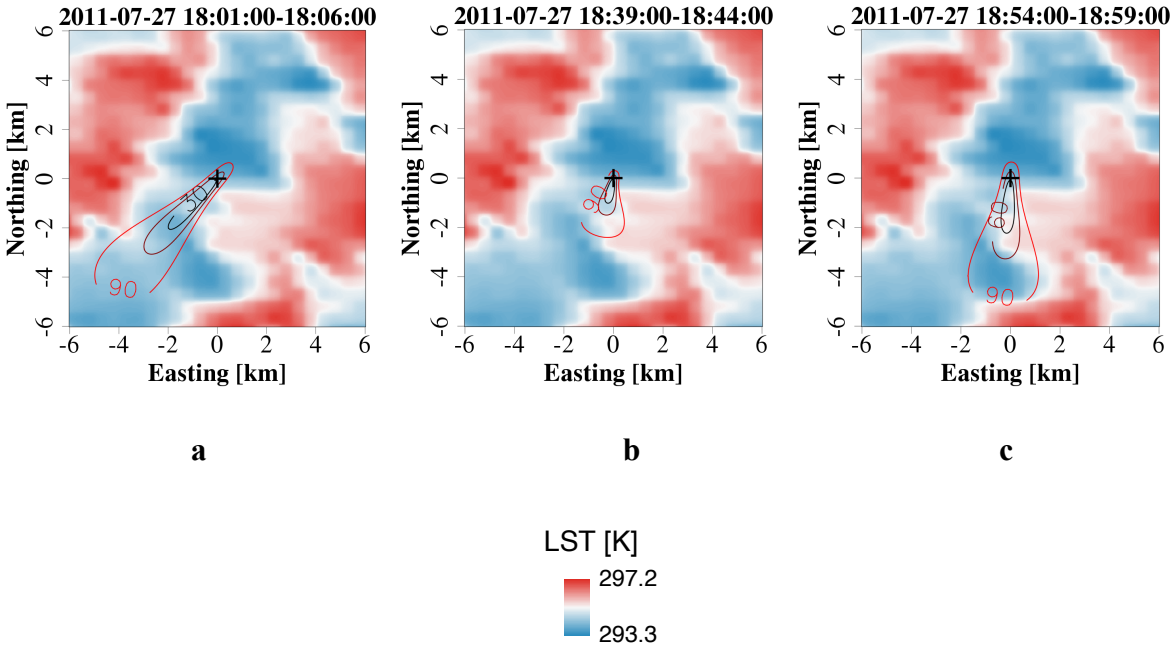


Figure 7 Flux footprint (30%, 60% and 90%, contour lines) for 122 m level measurements over MODIS-land surface temperature (LST) on July 27th, 2011, (a) 18:01–18:06, (b) 18:39–18:44, (c) 18:54–18:59 CST.

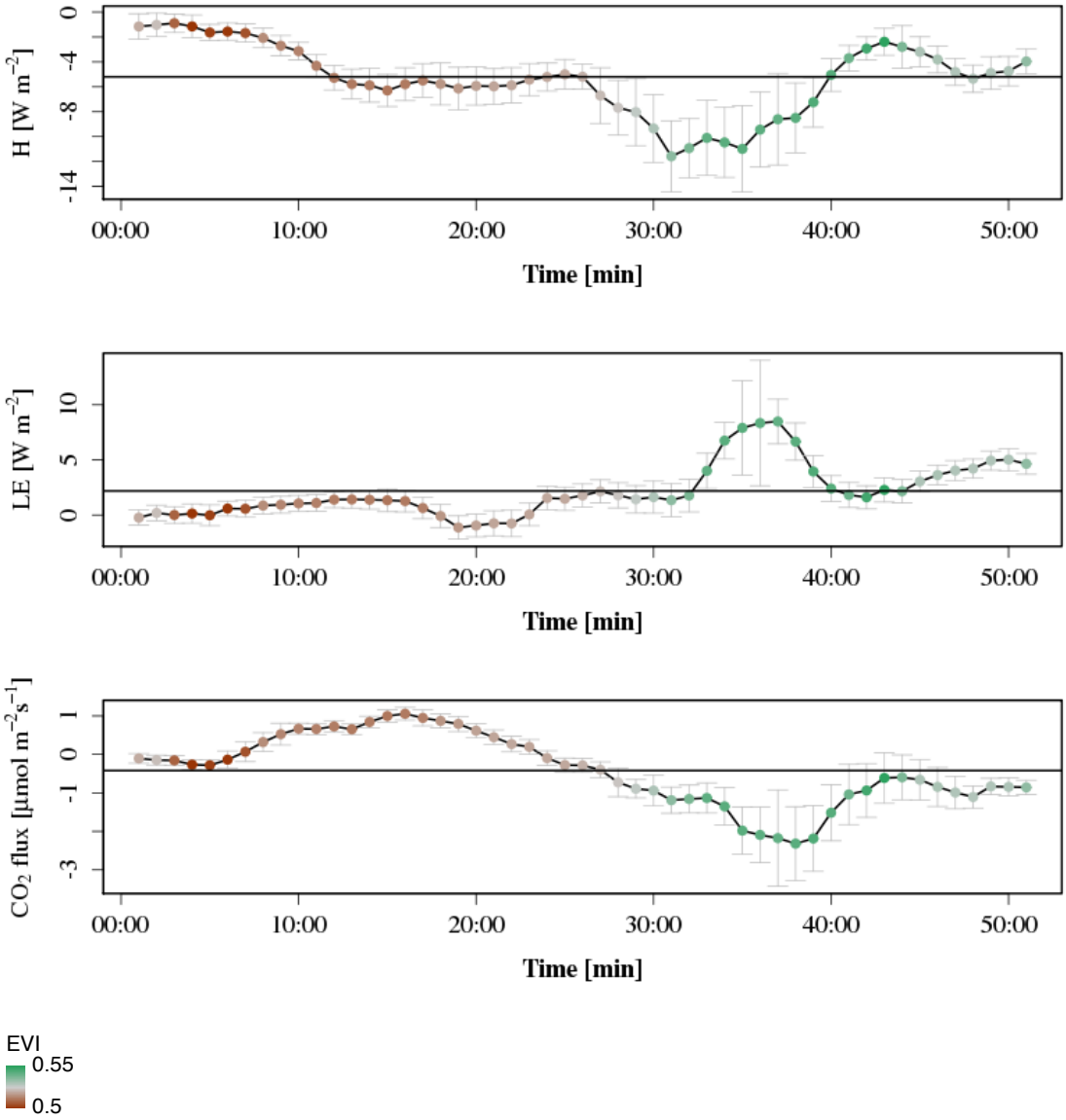


Figure 8 Time series of turbulent sensible heat (H), latent heat (LE), and CO_2 fluxes on July 27th, 2011, 18:00-19:00 CST. Colors and grey bars indicate the footprint-weighted EVI and 1 sigma random uncertainty of each flux observation, respectively. Black horizontal lines show the hourly-average fluxes 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.

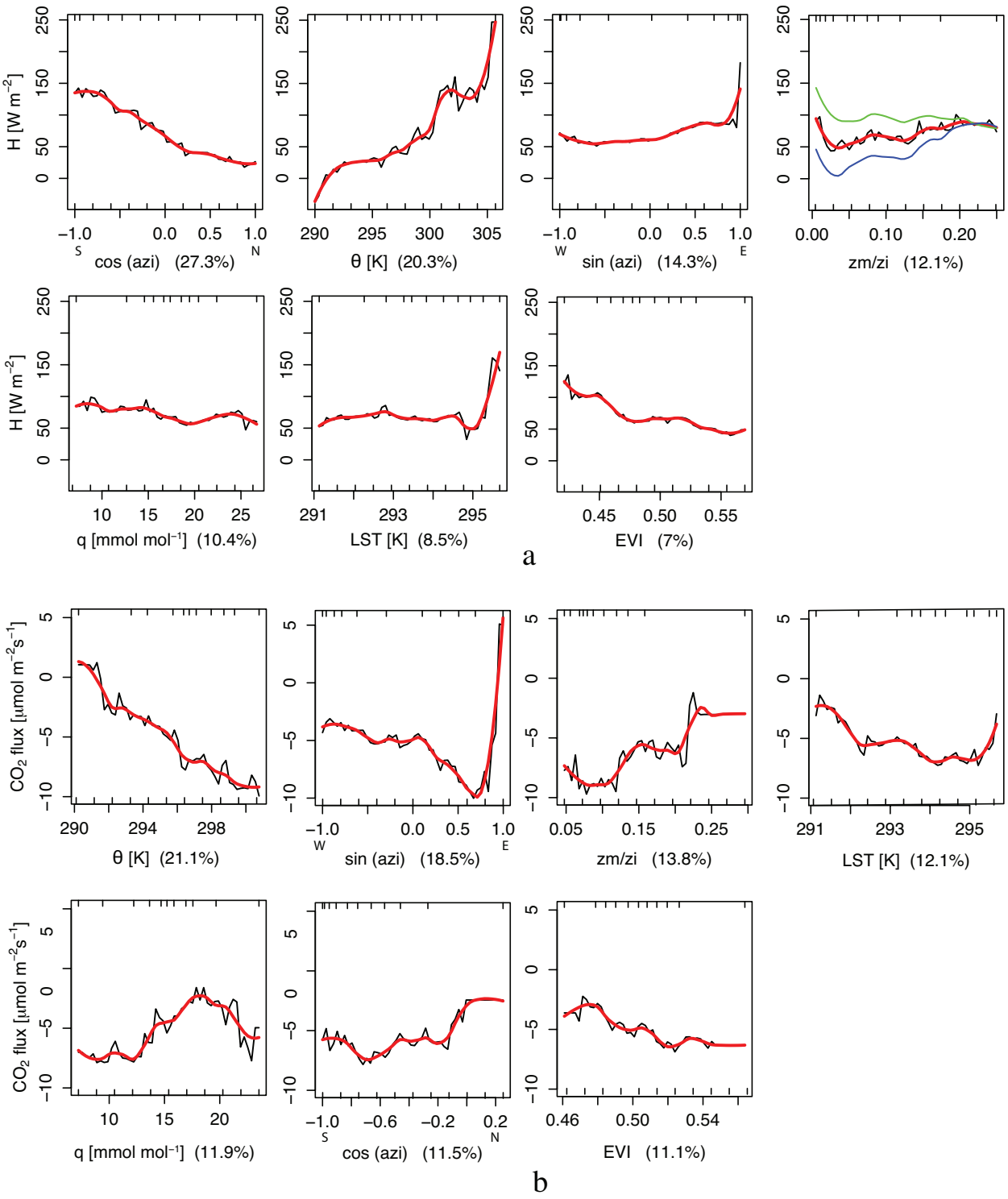


Figure 9 Equidistant response plots are shown for sensible heat (H, panel a) and CO_2 flux (panel b). The drivers are presented in decreasing order of explained variation in the flux signal (partial R^2 in braces). The black lines are the fitted integrated response over the range of one individual driver. 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.

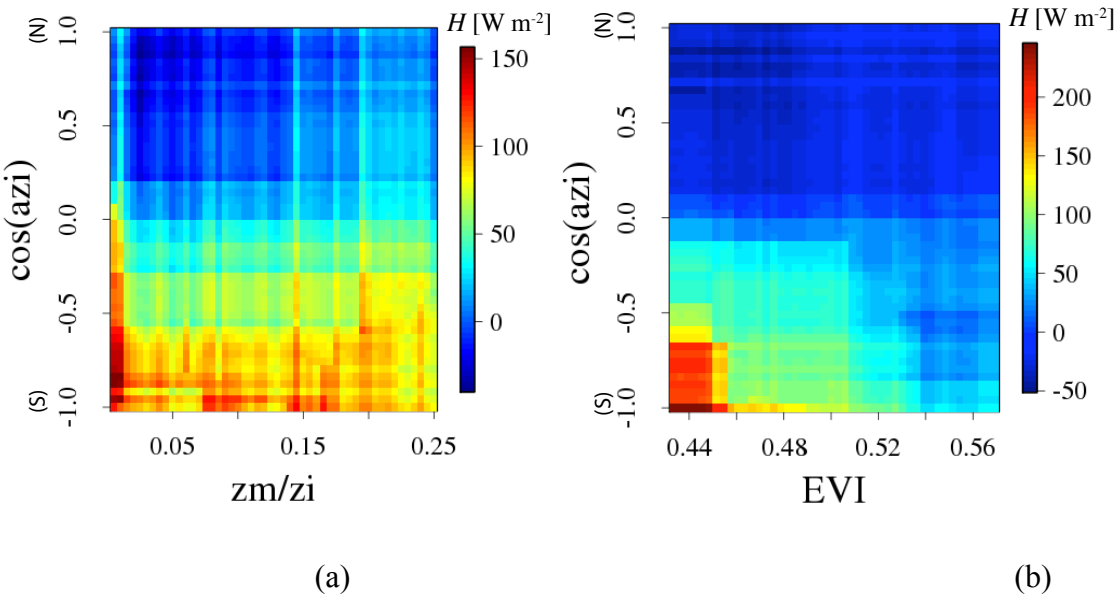


Figure 10 The response surface of sensible heat as function of (a) cos(azi) and z_m/z_i, and (b) cos(azi) and EVI.

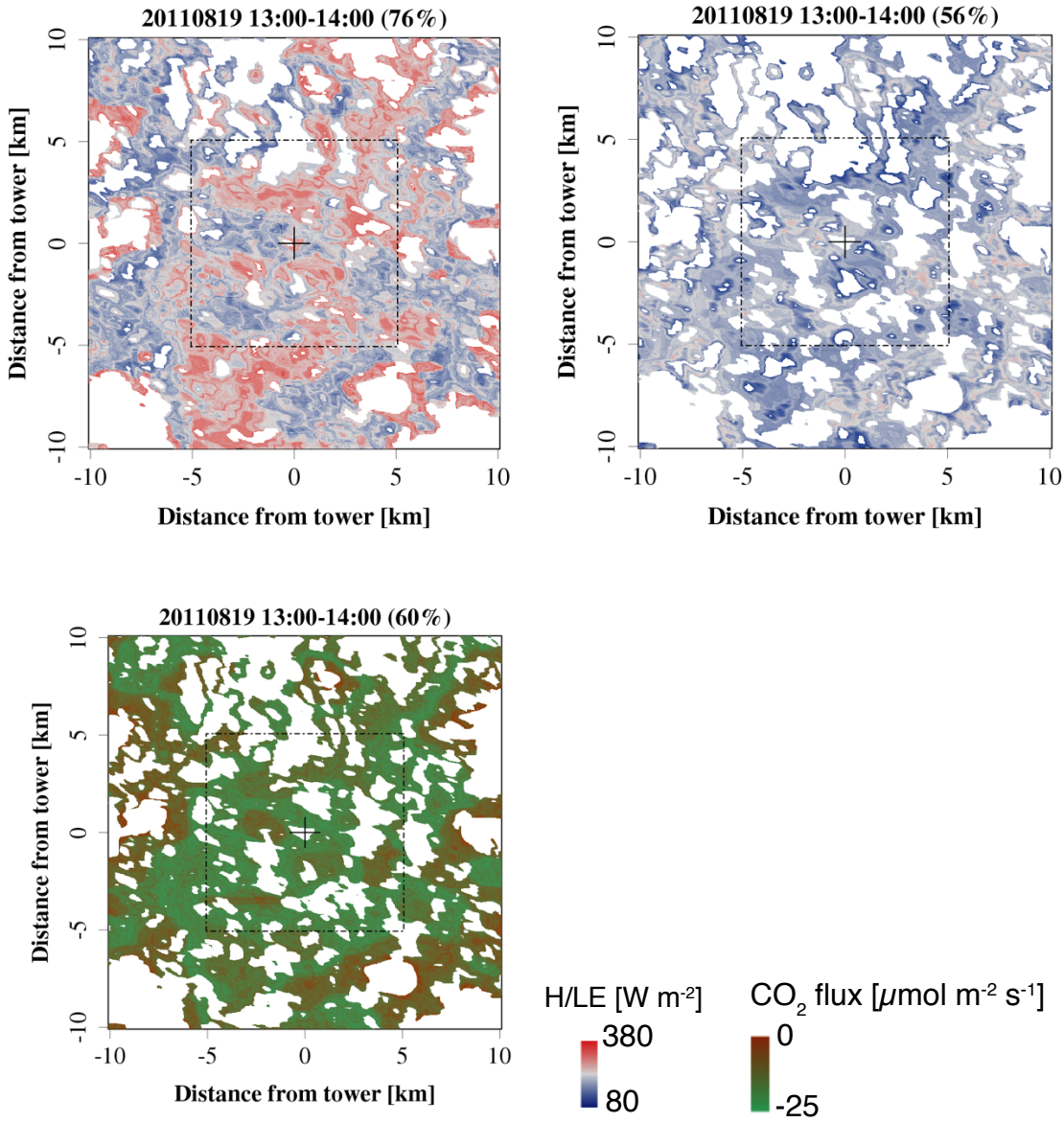


Figure 11 Averaged minutely projected flux grids of sensible heat (top left panel), latent heat (top right panel) and CO_2 (bottom left panel) over August 19th, 2011, 13:00-14:00 CST for tower-centered 20×20 km^2 target region. White areas are gaps that cannot be reproduced by ERF because their biophysical properties exceed the range of the training dataset.

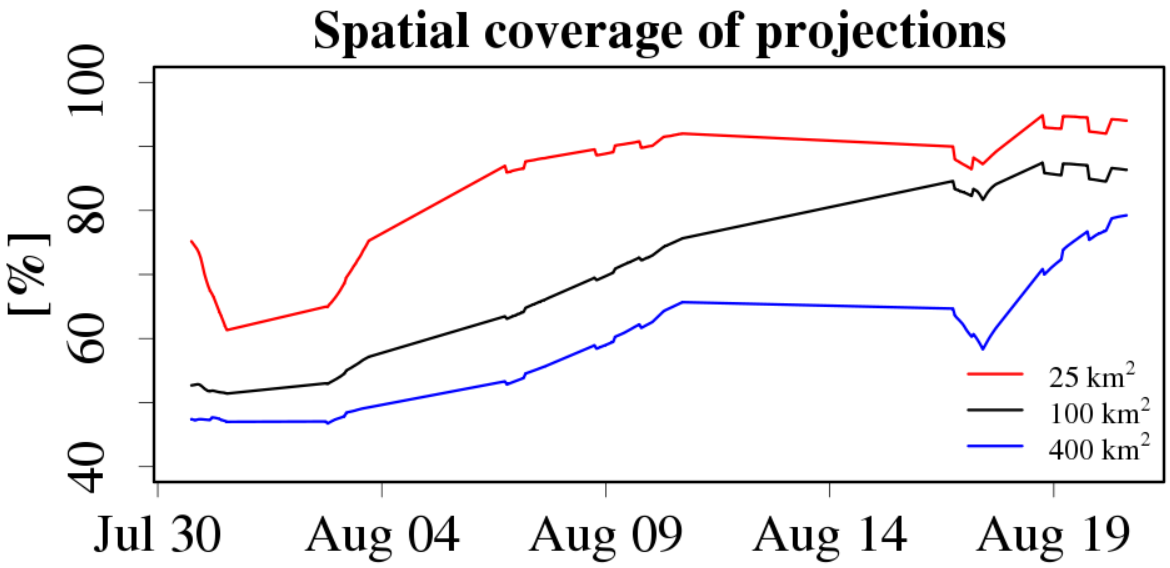


Figure 12 Time-varying spatial coverage of the sensible heat flux projection over 5×5 km^2 (red), 10×10 km^2 (black) and 20 km × 20 km area (blue).

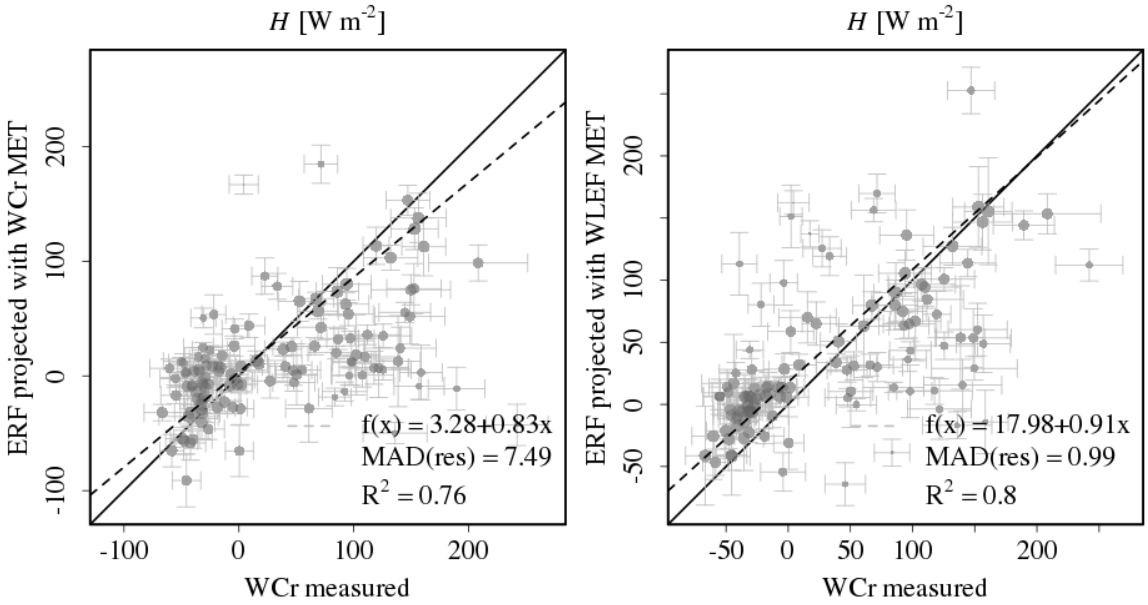


Figure 13 Scatterplot for Willow Creek (WCr) sensible heat flux measurements and ERF projections over 3 km × 3 km centered around WCr. Each point represents a one hour averaging period, and corresponding aggregation is applied to the minutely ERF projection with meteorological drivers measured at WCr (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.

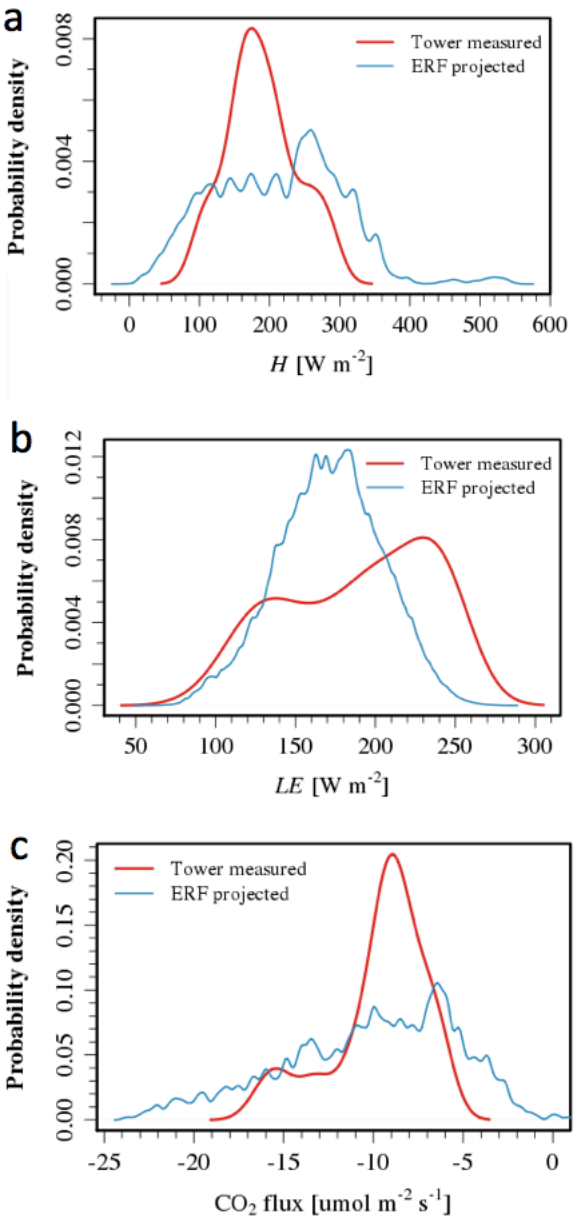


Figure 14 Probability density functions of sensible heat (H, panel a), latent heat (LE, panel b) and CO_2 (panel c) turbulent fluxes. Shown are minutely tower observations (red) and minutely ERF flux maps over 20×20 km^2 (blue) for August 19th, 2011, 13:00-14:00 CST.

3 Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance[2]

Keywords: Eddy-covariance; Mass continuity equation; Storage flux; Location bias; Energy balance closure; environmental response function

3.1 Introduction

The eddy-covariance technique has been used worldwide to monitor the surface atmosphere exchange for decades (Baldocchi, et al., 2001; Bonan et al., 2011). Theoretically, it provides reliable observations for comparison with remote sensing data and earth system models (ESMs) to constrain model performance and to improve our understanding of the impact of global climate change (Dietze et al., 2014; Richardson et al., 2012; Schaefer et al., 2012). However, despite its widespread use, the eddy covariance technique faces a number of challenges especially when attempting to monitor surface-atmosphere exchanges in conditions that deviate from assumptions used to derive the method.

Among those challenges, location bias (Desai et al., 2008; Chen et al, 2011; Schmid, 1997; Stoy et al., 2013) and energy imbalance (Leuning et al., 2012; Foken et al., 2011; Foken, 2008) are two major ones. Location bias refers to the surface influencing the eddy covariance flux varying temporally and usually within the areas of 10^{-1}–10^{1} km^2. This area is far smaller than typical ecosystem scales and the resolution of most earth system models (10^{2}–10^{4} km^2), leading to a mismatch of scale and potentially representativeness in model-data comparison. Furthermore, almost all flux towers suffer from an imbalanced energy budget: the sum of measured sensible

[2]This chapter is modified and excerpted with permission from Xu, K., Metzger, S., & Desai, A. R. (2017). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. *Agricultural and Forest Meteorology*. doi:10.1016/j.agrformet.2017.10.011.

and latent heat flux is consistently less than available net radiation minus ground heat flux, which may also indicate bias in other trace gas fluxes such as CO_2 (Foken, 2008).

Here, we argue that one root cause for these challenges are the inherent assumptions of the eddy covariance technique allowing the measured vertical turbulent flux (Figure 15b) to represent the net surface-atmosphere exchange (NSAE) across a volume (Figure 15a). This volume is confined vertically from the surface to the measurement height, across a horizontal target domain that is large in comparison to the flux footprint. The vertical turbulent flux measured at the top of the tower is assumed to be spatially representative across the entire target domain. However, in reality, the corresponding flux measurement footprint typically represents a small fraction (order 1%) of the target ecosystem or ESM grid cell.

Further, the location of this fraction changes with time. Many sites have differing frequency of wind directions and magnitudes at daytime versus nighttime (Xu et al., 2017). This results in vertical turbulent flux observations at daily scales being a convolution of the diurnal cycle with a related systematic change in flux footprint. In non-homogenous conditions, which are common, this temporally varying and small flux footprints directly leads to location bias in eddy covariance measurements (Xiao et al., 2014; Desai et al., 2015).

Second, long-term storage flux is often assumed to be either zero, filtered out using steady state criteria, or vertical profile measurements are assumed to be horizontally representative of the storage flux across the virtual box as a whole. However, in reality, the storage flux is not necessarily negligible for typical eddy flux averaging intervals, in particular for tall towers over complex surfaces. Lack of consideration of this term has been shown to contribute to the observed energy imbalance (Leuning et al., 2012). Moreover, each vertical profile observation used to calculate storage flux has its own footprint, resulting in an influence area that increases

with measurement height. This can be another source of location bias in flux tower observations (Schmid 1997; Raupach et al., 1988).

Third, advection terms are typically assumed to cancel each other or to be smaller than the turbulent flux. However, local advection can contain spurious low-frequency mesoscale contributions to the NSAE, which has been argued as another source for energy imbalance (Finnigan, 2003; Kanda et al., 2004; Foken 2008; Eder et al., 2015; Sakai et al., 2001). Low-frequency flux contributions can be caused by turbulent organized structures (TOS; Kanda et al., 2004; Finnigan 2003) and/or secondary circulations associated with surface heterogeneity (Schlegel et al., 2014; Eder et al., 2015). In the presence of these structures, strong convection of warmer/wetter air occurs in spatially confined updraft zones, countered by a slight subsidence of cooler/dryer air across the majority of the target area. The corresponding timescales exceed typical eddy covariance averaging periods, thus these structures manifest themselves as advection rather than as turbulent contribution to the net surface-atmosphere exchange. In addition, this flux contribution can appear as a component of horizontal advection, vertical advection and storage, or combinations thereof, depending on the location of updraft and downdraft zones with respect to the measurement location. Due to the skewed distribution of updraft and downdraft zones, heat fluxes will thus tend to be underestimated when only considering the turbulent vertical flux in the standard eddy covariance approach (Mauder et al., 2008).

To mitigate the first assumption (representativeness) on vertical turbulent flux, two main upscaling approaches, process-based and data-driven approaches, have been utilized. However, each is 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. Purely data-driven scaling (Xiao

et al., 2014, 2008; 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. More recently, Metzger et al., (2013) and Xu et al., (2017) developed the environmental response function (ERF) approach that combines the process-based and data-driven approaches. The underlying principle of ERF is to relate high-frequency (minute-to-minute) fluxes over fast-varying footprints with appropriate spatial or temporal drivers, e.g. meteorological and surface ecological forcings. The extracted relationships are then used for spatio-temporal mapping over a large domain.

To address the second and third assumption, Metzger (2017) developed ERF further, and derived the ERF virtual control volume (ERF-VCV) framework. In essence, ERF-VCV attempts to apply the ERF technique to all terms in the continuity equation, including storage flux, advection and turbulent flux, to estimate the control volume net surface atmosphere exchange (NSAE). The two main assumptions of ERF-VCV theory are: i) the eddy covariance observations are dominated by surface flux, not entrainment, and ii) all relevant atmospheric and surface dynamics / state-space combinations are sampled by the eddy covariance platform.

Here we ask, based on an example case using one single tower during July and August 2014 over a heterogeneous environment of AmeriFlux Park Falls WLEF very tall tower in North Wisconsin, USA:

1. Can assumptions made in ERF-VCV theory be fulfilled in a real-world setting and enable the method to retrieve VCV-estimated storage and vertical heat fluxes?
2. How do ERF-VCV produced heat fluxes compare in magnitude and pattern to the tower-measured turbulent heat fluxes and what does it imply for location bias?
3. How does ERF-VCV enable addressing advective heat fluxes?

4. Do the virtual control volume integrated heat fluxes provide insight into how energy balance closure can be further addressed?

To answer these questions, we first introduce the climate, biophysical properties in the study area and footprint composition of the WLEF tower data (Sect. 2.2.1). The methodology of ERF-VCV and associated uncertainty algorithms are described in Sect. 2.2.2. We present the extracted relationships, scaled storage and vertical flux grids and the associated uncertainty budget in Sect. 3.3. The previous four key questions are discussed in Sect. 3.4, and conclusions are provided in Sect. 3.5.

3.2 Materials and methods

3.2.1 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, within the Park Falls Ranger District of the Chequamegon-Nicolet National Forest, Wisconsin, USA (Xu et al., 2017). The surrounding landscape is a mix of wetlands and upland forests. The tower footprint climatology samples a landscape that is representative of much of the Upper Midwest U.S. forested region (Desai et al., 2008, 2015). 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. 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 variables, storage, turbulent, and vertical advection fluxes, as well as remote sensing products. Tower-based 10 Hz observations were chosen from 6th July to 31st August 2014 for WLEF at both 30 m and 122 m levels. The flux footprint along-wind distance was 0.1–0.8 km and 2–5 km for 30 and

122 m height measurement at 90% cumulative level (Figure 16), respectively.

For 30 m and 122 m levels, fast response wind speed and air temperature were derived from a sonic anemometer (Applied Technogies., Inc. Seattle, USA, ATI Type K). Dry mole fraction of water vapor were measured by a closed-path infrared gas analyzer (LI-COR, Inc. Lincoln, USA, LI-6262) at both levels. 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).

Land surface temperature (LST) and enhanced vegetation index (EVI) were chosen as biophysical surface drivers. These two 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) LST. 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).

3.2.2 Environmental response function - virtual control volume (ERF-VCV) approach

The ERF flux scaling procedure for tower eddy covariance measurements is based on Metzger et al., (2013) and Xu et al., (2017). The underlying principle of ERF is to relate high-frequency (minute-to-minute) fluxes via fast-varying footprint estimates with appropriate spatial or temporal drivers, e.g. biophysical surface and meteorological forcings. The extracted relationships are then utilized for spatio-temporal mapping over a domain that exceeds the typical footprint extent. The ERF method uses the footprint variation to re-assemble the NSAE. In order to project flux into one area, the exact same area doesn't necessarily have to be

measured as long as its properties are within the state space of the training dataset used by the machine learning.

Unlike the work of Xu et al (2017), where ERF was applied solely to vertical turbulent flux for a single month, here ERF flux scaling is realized not only for vertical turbulent flux ($\overline{w'c'}$), but also for low-frequency turbulent contributions to the vertical advection term ($\bar{w}\bar{c}$) and storage flux ($\frac{\partial \bar{c}}{\partial t}$) observations over a longer-time period (2 months) to retrieve flux across the whole virtual control volume (VCV). This procedure for the first-time permits addressing these terms in the mass continuity equation that are neglected by the standard eddy covariance methodology.

We first determined appropriate temporal scales for vertical transport, and calculated high-rate (one-minute) flux responses using wavelet discretization for tower-based vertical transport permitting inclusion of transporting scale up to three hours (Sect. 3.2.2.1). Next, the storage flux was determined using measurements at multiple vertical levels (Sect. 3.2.2.2). Lastly, ERFs were extracted using machine learning (Figure 4) and used for projection for both vertical and storage flux (Figure 19, Figure 20) in Sect. 3.2.2.3. Our routines were developed in GNU R version 3.1 (R Development Core Team, 2014), and code and examples are being developed for a public repository (Metzger et al., 2017).

3.2.2.1 Wavelet discretized vertical exchange

Building on Metzger et al. (2013) and Xu et al. (2017), several preprocessing steps were performed: de-spiking after Brock (1986) (Table 1, Row 1) and Starkenburg et al.(2015), planar fit rotation (Wilczak et al., 2001), fixed lag correction using maximum correlation, and point-by-point conversion of sonic temperature to air temperature based on Schotanus et al. (1983). Only hours with more than 80% available raw data were used for further analysis (Table 1, Row 2).

Since flux footprint varies rapidly even within one hour (Xu et al., 2017), we use high-frequency (sub-hourly) flux responses to avoid footprint blending. At the same time, the spectral range of aggregated transporting scale from 10 Hz up to three hours permits combining high-frequency vertical turbulent flux and vertical advection (up to three hours) together into a single "vertical flux" term (Finnigan 2003, Steinfeld et al., 2007, Mauder et al., 2008, Mahrt 1998). Wavelet decomposition (Torrence and Compo et al., 1998) can satisfy these two requirements to achieve high temporal resolution while including transporting scales up to several hours (Charuchittipan et al., 2014). However, wavelet decomposition assumes data are circulate, this assumption leads to larger uncertainty at the begin and end times of the dataset compared to its center (Torrence and Compo, 1998; Metzger et al., 2013). Therefore, the vertical flux over transporting scale of up to three hours were calculated using 15 hours of raw data in order to avoid this edge effect. Here, we integrated over wavelet transport scales of up to three hours, with flux results discretized over a five-minute window that for each observation moves one-minute forward in time. Within the entire observation period, 55,160 one-minute vertical flux observations were obtained and qualified for machine learning.

3.2.2.2 Storage flux

We also determined the averaging time for storage flux. Following Finnigan, (2006), we should avoid storage flux estimates influenced by single or a small number of eddies influencing the observations. Instead, we should use a period long enough to capture an adequate ensemble of these eddies. Here we consider the integral time scale of the turbulent time series between these eddies, and multiply it by one order of magnitude as the basis of storage flux computation. Considering the maximum integral space scale (98.4 m) and the average mean wind speeds of 2.2 m s^{-1} and 0.5 m s^{-1} at 122m and 30 m, respectively, we determine 15 minutes as storage flux averaging time. For example, the storage flux estimates for timestamp 08:00:00 is calculated as

the time average of measurements from 8:00:00 to 8:14:59 minus the average from 7:45:00 to 7:59:59. Within the entire observation period, 3,304 storage flux observations are obtained and qualified for machine learning.

3.2.2.3 Environmental response function extraction and projection

The operator underlying the extraction of environmental relationships is the flux footprint model which links the flux responses to surface biophysical drivers. ERF builds the relationships among the observed fluxes, meteorology and footprint weighted surface properties to unveil the whole, time-varying flux field. In addition, by including transporting scales of three hours, one tower may sample updraft branch and downdraft branch in a mesoscale circulation, which principally enables ERF to reproduce large eddy circulations over the virtual box (Figure 17).

Storage flux is the time-rate-of-change of a state variable such as temperature or mixing ratio of water vapor, as opposed to the state variable itself. Thus, we need to also consult a flux footprint model instead of concentration footprint model. Therefore, storage flux environmental relationships were also generated with observed storage exchange (response) and flux footprint-weighted biophysical surface properties and meteorological forcings (drivers). Flux footprint matrixes were calculated (Metzger et al., 2012; Kljun et al., 2004). The footprint model is valid for certain ranges of input parameters. Instead of discarding footprints when the input parameters that exceed the allowable range, we instead set parameters that exceeds the input range to be minimum or maximum, i.e. roughness length less than $1e^{-5}$ is set to be $1e^{-5}$, and larger than 1 is set to be 1, vertical wind speed variation 0.23–1.23, and u* less than 0.2 is set to be 0.2, only for the purpose of footprint matrix calculation. Only for the purpose of footprint modeling, 11.7% of the data were constrained to match the valid range of footprint model inputs. Both storage and vertical fluxes are filtered for (i) unreliable nighttime data; (ii) spikes; and (iii)

connectivity with the surface using the integral turbulence characteristic (ITC) test after Foken (2008), as shown in Table 1, Row 3–5, separately. It should be mentioned that nighttime data considered as unreliable for the purpose of this study may still contain valid information for dealing with different phenomena.

In terms of driver selection, solar azimuth angle azi, relative measurement height within boundary layer, temperature and water vapor gradients between surface and atmosphere were selected according to Xu et al., 2017. To explain the diurnal cycle and solar radiation, we derived and included cos(azi) and sin(azi). Considering the relative measurement height in the ABL, z_m/z_i, not only permits combining eddy covariance measurements from multiple heights z_m, but also to explicitly account for vertical flux divergence during ERF projection. This property allows us to project to all vertical levels above displacement height. Lastly, the vertical gradients of temperature and water vapor were explained using air potential temperature (θ in K) and mole fraction of water vapor in dry air (q in $mmol\ mol^{-1}$), as well as LST and EVI as corresponding land surface drivers. LST and EVI matrixes were downscaled from MODIS data products, bi-linearly to 100 m in space, and linearly to one-hour in time. Surface properties are more responsible for spatial variability of the response in machine learning, while meteorological drivers, e.g. air temperature, are more responsible for the diurnal cycle and temporal variability. Considering the coarse temporal resolution (8/16 day) of surface properties, we chose to use simple linear interpolation in time for LST and EVI matrixes.

Building on Metzger et al., (2013) and Xu et al., (2017), we used boosted regression trees (BRT) as machine learning technique, which is based on categorization and regression. 55,160 vertical fluxes and 3,304 storage fluxes served as the training dataset in machine learning to produce fluxes over grids and volumes, respectively. Ten cross-validations were operated at 10 nodes before the model with the best agreement between the fitted fluxes and the training dataset was

selected. In the end, 70,000 split points were used for vertical flux regression and 3040 split points for storage flux regression.

The extracted ERFs were summarized in equidistant response-sensitivity plots (Cacuci, 2003), which show the driver-response relationships stored in the ERF (Figure 4). 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 the explained variance was determined from the reduction of R^2 when removing one driver at a time. The resulting ERF acted as a transfer function and was applied to project vertical flux to each 100 m grid cell across the 20×20 km^2 target domain at 122 m, and to project storage flux to the whole virtual box at five levels, 30 m, 53 m, 76 m, 99 m and 122 m, at hourly interval. During projection, for the meteorological drivers we used the median value during the time interval, assuming that the atmospheric state above the target area was spatially homogeneous. In particular, above the blending height, this assumption is weak compared to invoking a homogeneous land surface in the standard eddy covariance technique.

When summarized over the study period, due to 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 the mean of the monthly-mean diurnal cycle.

3.2.2.4 Uncertainty budget

To evaluate the significance of the presented approach, we estimated the uncertainty budget for both storage and vertical fluxes. We used stratified cross-validation to evaluate how well ERF-VCV performs when projecting to areas the tower footprint had never covered during the training period. Following Xu et al., (2017), we divided the target area into four quadrants:

northeastern, southeastern, southwestern and northwestern. 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 ERF along with the state variables from the omitted quadrant were used for projection; (iii) The resulting projection was compared to the observation. In this cross-validation, all uncertainty sources through the ERF-VCV method, i.e. input state variables, footprint modeling, and machine learning, were included in this uncertainty quantification.

We use median and median absolute deviation (MAD) for quantifying systematic and random uncertainty, respectively (Croux and Rousseeuw, 1992; Rousseeuw and Verboven, 2002). The resulting combined uncertainty estimates correspond to a single projected grid cell in the for virtual box.

3.3 Results

3.3.1 ERF-VCV projected storage flux

Figure 4 shows ERFs extracted from BRT. To note, the absolute values shown on the y-axis of Figure 4 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. However, these equidistant plots are most useful for revealing the relationships among driver and response stored in the ERF. Temperature time-rate of change shows a strong diurnal cycle, negative during daytime and positive during nighttime. The pulse at sin(azi) = 0.4 corresponding to 7–9 a.m. local time is the sign of increase in storage flux when convective eddies are overcoming nighttime stable stratification during initiation of the convective boundary layer (CBL). Temperature time-rate of change is larger in shallower ABLs (larger zm/zi), and also has a positive relationship with LST and θ, while it has a negative

relationship with EVI. The ERF-VCV predictions fit very well the observed fluxes (-1% for vertical H, 0% for vertical LE, -4% for dT/dt, and -8% for dq/dt), but in all cases the 99% confidence intervals include the unity slope.

On Aug 17, 2014, the lower part of the convective layer developed as follows: From 7:00–8:00 a.m., an initiation of buoyancy on the surface due to solar forcing was detected (Figure 19). Warm bubbles built up and began to release from the surface. During 8:00–9:00 a.m., critical buoyancy was reached. As a result, heated air detached from the surface in discrete events, leading to the creation of a spatial pattern with distinct zones that exhibit varying degrees of heating. From 9:00 a.m.–10:00 p.m., the whole volume was continued to be heated.

3.3.2 ERF-VCV projected vertical flux

Figure 20 shows an example of the domain-scaled vertical transport of H flux over the predefined 20 km × 20 km target area at 86 m and the measurement height, 122 m. The white cells in the figure are areas with state-space combinations of drivers for which no extracted response relationship exists. Over the whole experiment, the coverage was extended from the original 1% (average footprint area relative to 20 km x 20 km) to 92%±3%, and 94% ± 3% for 20×20 km^2 target domain for H, LE respectively, where the tolerance here is one standard deviation.

In Figure 20, for Aug 19, 2014 noontime, the sensible heat flux at 76 m was positive over warmer surfaces and negative over cooler surfaces. This may indicate suitable conditions for surface heterogeneity-induced mesoscale circulations. In addition, vertical flux divergence and blending can be quantified explicitly in space: Buoyant eddies were emitted from individual surface patches giving the impression of a clear spatial separation at lower levels, e.g. 76 m. The flux at 122 m level was spatially less distinct as a result of blending (Figure 20b).

3.3.3 Volume controlled net surface atmosphere exchange

When superimposing vertically integrated projected storage flux with the vertical flux grids at 122 m, ERF-VCV NSAE of H and LE were 32.3 Wm^{-2} and 74.4 Wm^{-2}, 33.8% and 15.5% larger compared to tower observed turbulent flux, 24.1 W m^{-2} and 64.4 Wm^{-2} for the study time period. Here, vertical flux refers to the sum of vertical turbulent flux and vertical advection flux with transporting scale up to three hours. The sum of NSAE of sensible and latent heat flux was 106.7 Wm^{-2}, 20.6% greater than the turbulent observation, 88.5 Wm^{-2}. The standard deviation of the diurnal cycle over the study period was 120.9 Wm^{-2} and 228.9 Wm^{-2} for ERF-VCV-projected H and LE, and 149.5 Wm^{-2}, 256.6 Wm^{-2} for observed turbulent H and LE, because spatial averaging ERF-VCV projections were less scattered compared to direct observations. The fitted linear relationship shows that the ERF-VCV projections had a larger daytime-nighttime amplitude (Figure 21). Temporal standard deviation of the spatially aggregated flux first decreased and then stabilized at about 10–15 km spatial scale (Figure 22). Incorporation of the ERF-VCV approach also reflects the estimated energy fluxes increased until converging around the same 10–15 km spatial scale.

Storage flux is non-negligible even at the monthly scale, and exhibits a sizeable diurnal cycle at hourly temporal resolution for the virtual box of 122 m height, e.g. 42 Wm^{-2} amplitude (~30% of vertical transport flux) for storage flux of H (Figure 23).

The median systematic and random uncertainty terms (median absolute deviation, in parenthesis) per single projected cell are 10% (162%), -19% (185%), -15% (232%), and -19% (178%) for vertical flux of H, storage flux of H, vertical flux of LE, and storage flux of LE, respectively. Here, we see that the ERF-VCV approach tends to overestimate vertical H by 10%, and to underestimate vertical LE by 15%. The ERF-VCV approach underestimated the storage fluxes of H and LE by 19%. On this basis, an overall uncertainty of < 15% for H and LE vertical

fluxes and < 20% for storage flux is determined. To note, this systematic uncertainty is raw output from the machine projection, and subsequently corrected across the target-area via site-specific linear regression. In result, the +20.6% or +18.2 Wm^{-2} difference for the sum of H and LE between ERF-VCV-controlled NSAE projection and turbulent flux tower observation is shown. In addition, ensemble random uncertainty becomes very small when aggregating flux grid cells over the whole target domain: the ensemble random uncertainty for hourly projection over the target domain is confined to within 1% for all fluxes.

3.4 Discussion

3.4.1 Evaluation of the assumptions made in the ERF-VCV theory

A number of assumptions influence the reliability of the ERF-VCV approach. The first assumption is that the tower sufficiently samples both updrafts and downdrafts. We believe this is satisfied in the current case study because the near-zero averaged vertical wind speed ($\overline{w} = 0.01\, ms^{-1}$) over the entire study period and non-zero hourly $\overline{w}$. These two characteristics imply that both updrafts and downdrafts were sampled by the WLEF tower during the study period.

The second assumption is that observations are dominated by surface fluxes rather than entrainment from the top of the boundary layer, which ensures the relationships reflect information content from the chosen surface drivers. This assumption is fulfilled by filtering flux responses for connectivity with the surface through turbulent mixing tests, primarily the ITC test, which tends to omit data affected by strong negative vertical velocity. The ITC test does not omit situations when surface heterogeneity-induced downdraft occurs in a growing boundary layer, since the net velocity is still upward. Thus, the downward branch of a mesoscale circulation is still sampled by the tower and used as training data in ERF-VCV.

Another implicit assumption is that flux footprint can attribute the main contribution of vertical flux to appropriate surface area even with transporting scale up to three hours. In other words, we have to assume that the majority contribution is from the last touchdown with the surface instead of entrainment (Flesch et al., 1996; Metzger, 2017). Since the majority contribution of the vertical flux is from vertical turbulent flux term (Finnigan 2003, 2004), we use a flux footprint model instead of concentration footprint model. Further, the extracted relationships are built between flux and flux footprint-weighted land surface drivers as well as meteorological drivers. Although flux footprint may fail to attribute low frequency flux contribution to the appropriate surface drivers, the response relationship can still be propagated via the corresponding combination of meteorological drivers. At this study site, the extracted ERFs and flux projections appear sound. This suggests that over the ensemble of thousands of observations the simple footprint parameterization is sufficient to accurately and precisely relate a large fraction of the flux responses to their surface drivers.

Though these assumptions do require additional testing, our results support the idea that ERF-VCV projections enable explicitly identifying and quantifying vertical turbulent flux and storage flux. As shown in Figure 20, the method was able to detect the initiation of buoyancy from solar forcing, development of critical buoyancy leading to a well-mixed volume, and continue mixing driving by mesoscale circulations generated by surface heterogeneity. These results are remarkably similar to theoretical expectations of boundary-layer mixing.

The ERF-VCV projections suggest that the ERF-VCV technique is a practical and theoretically sound approach to retrieve storage and vertical transport flux in a target box with a maximum (correctable on site-level) systematic uncertainty of less than 20%. The estimation of these terms and further evaluation of them allow us to directly examine several unmet assumptions of eddy covariance in real-world settings which we discuss below: i) location bias of vertical

turbulent flux; b) difference of profile to whole domain storage flux; and c) the effect of neglecting vertical advection.

For future development, ERF-VCV requires development of a transporting-scale footprint model to identify and quantify large transporting scale eddies. The transporting scale should not be scaled beyond the spectral gap to avoid ambiguous interpretation of energy transport at a different scale, e.g. synoptic system. Storage flux with inclusion of transporting scale up to three hours could be obtained using wavelet decomposition to match the transporting scale of vertical flux. The overlapping information content within different drivers (multicollinearity) can lead to the extraction of process relationships by the machine-learning that could not be reconciled with physical theory. It will be interesting to work on reconciling additional constraints and multicollinearity in machine-learning. Lastly, with sub-canopy eddy covariance measurements, ERF-VCV should be applicable to capture the features of eddy scalar and heat transport in sub-canopy.

3.4.2 The implication on location bias in eddy covariance measurements

Combining reliable, precise measurements with time-frequency decomposition and flux footprint modelling allows us to transform our observations into a unified Eulerian coordinate representation, through which we can finally estimate terms that used to be inaccessible. Our study reflects that storage flux has very strong diurnal amplitude (42 W m^{-2}). Therefore, storage flux should not be ignored in observation and model-data comparison, at hourly temporal resolution, especially in the case of tall towers. However, in many cases storage flux is usually assumed to rapidly converge to zero and thus is not measured at many sites (Foken 2008).

ERF-VCV estimation has a larger vertical H than tower turbulent observations during nighttime, likely owing to a higher LST across the target domain than the footprint area (Figure 16).

Similarly, a smaller vertical LE during daytime is observed in association with a slightly lower EVI across the domain. In addition, ERF-VCV NSAE shows larger magnitude both in daytime and nighttime despite substantial scatter in the tower observations. This reflects the influence of the diurnal cycle on the transient footprint. Further, in theory, ERF-VCV-projected vertical flux is expected to be larger than tower turbulent observation because projected vertical flux includes the energy contribution from mesoscale eddies (Foken et al., 2006). The inclusion of mesoscale eddies tends to increase emitted heat flux, since spatially confined thus under-sampled updrafts are net warmer compared to spatially expansive thus over-sampled downdrafts. The resulted difference between the tower observation and ERF-VCV estimation reflects that WLEF tower cannot consistently represent the mean VCV estimated surface-atmosphere flux, implying location bias is pervasive in this case and potentially reflective also of other flux tower sites in heterogeneous terrain.

3.4.3 The role of horizontal advection and the implication on spatial transporting scale

Theoretically, in a mesoscale circulation we expect vertical advection to be compensated by horizontal advection and storage flux (Mahrt 1998, Foken 2003, Mauder 2008). Hence, studies considering only vertical advection instead of advection from both vertical and horizontal directions have incurred criticism. For example, Finnigan (2003) and Mauder et al., (2008) argue that only considering the vertical advection term can produce unrealistic large convection and rather noisy results on an observation-by-observation basis.

Here, we argue that the spatial aggregation of projected fluxes over heterogeneous terrain improves the resilience of the ERF-VCV projection to horizontal advection. Especially when the spatial scale of aggregation is large, the net low-frequency vertical flux due to updrafts and downdrafts trends towards its ensemble mean. As shown in Figure 22, the temporal standard deviation of the spatially aggregated flux first decreased and then stabilized at about 10–15 km

horizontal scale. This indicates that ERF-VCV also projects the localized vertical advection resulting from the observation-by-observation compensation of horizontal advection. However, when integrated over a critical landscape scale, positive and negative compensatory fluxes cancel out each other, and only the net low-frequency flux remains. This justifies the choice of neglecting horizontal advection when aggregating over a horizontal scale of order 10 km as compared to point-by-point observations. Our finding here generally agrees with previous studies suggesting that w = 0 over a large spatial domain is a much weaker assumption compared to w = 0 for an individual measurement location (Mahrt 1998).

In addition, the indicated surface patch size of 15 km implies that it is the minimum reliable patch size in the comparison of WLEF eddy covariance flux observations with model and remote sensing data, akin to a critical landscape scale. Our approach thus provides a suitable bridge for the spatial gap between WLEF tower-measured fluxes and both, remote sensing products and ESM outputs.

3.4.4 Implications of ERF-VCV on the energy balance closure problem

As discussed in Sect. 3.4.1, surface heterogeneity-induced updrafts and downdrafts were sampled by the tower at different times during the study period **(**Figure 17**)**. While changing wind direction and varying source areas complicate direct interpretation of the flux observations, they provide an opportunity for including secondary circulations in ERF-VCV.

At any given time a single tower is unable capture both updrafts and downdrafts. Due to the skewed spatial distribution of updrafts and downdrafts, a tower observes sparser but stronger updrafts only occasionally, if at all (Figure 17b). More frequently, a tower observes abundant but weaker downdrafts (Figure 17c), or cannot capture any low frequency energy transport at all (Figure 17a). For example Kanda et al. (2004) and Mahrt et al. (1998) link this spatial

patterning and corresponding conditional sampling characteristics to a buoyancy-related underestimation in particular of the sensible heat flux.

Different from the standard tower eddy covariance technique, ERF-VCV provides a potential approach for improving energy balance closure through spatially and temporally explicit flux projections: the projections include the likely spatial distribution of both low frequency updrafts and downdrafts. Once aggregated to the critical landscape spatial scale the compensatory fluxes cancel out. The resulting net flux trends positively to its ensemble mean (Figure 22 c,d), which exceeded the standard tower flux observations over the study period. This warrants investigating further impacts on the energy balance closure problem, once spatially explicit net radiation and ground heat flux are available, e.g. from intensive in-situ observations.

3.5 Conclusions

Using AmeriFlux Park Falls WLEF tall tower in North Wisconsin, USA during July and August 2014, ERF-VCV proved useful for retrieving the volume-controlled net surface atmosphere exchange (NSAE). This retrieval is achieved by resolving the storage flux, vertical turbulent, and vertical advection fluxes, which are not easily measured. ERF-VCV can improve the tower observation from a footprint-variable representation to a fixed-coordinate representation. This aids reducing the location bias typically incurred from single-location vertical turbulent flux and single-profile storage flux measurements, as well as the influence of vertical advection. Particularly, in this study, storage flux did not converge to zero as often assumed and had significant diurnal cycle, and should thus not be ignored when comparing eddy covariance measurements with earth system models (ESMs) at hourly or finer temporal resolution. Low frequency flux contributions were detected in this study, and inclusion in the ERF-VCV landscape-scale exchange of sensible and latent heat led to a 20.6% increase over the tower observations. In addition, the derived flux can spatiotemporally resolve mesoscale circulation

that contribute to this source of energy imbalance. Further, we show that spatial aggregated fluxes over heterogeneous land cover resilient to horizontal advection. Lastly, in our case we find that ERF-VCV-estimated NSAE always increased the turbulent heat fluxes and thus provides a promising research direction for improving energy balance closure.

Substantial improvements are still possible and needed for these kinds of scaling and rectification methods. First, high intensity in-situ observations or large eddy simulations (LES) model results can be used to comprehensively evaluate and verify assumptions of the ERF-VCV process. Second, with spatially explicit net radiation and ground heat flux, the potential for improving the energy balance closure problem can be studied. Third, multiple flux tower or airborne eddy covariance data lend themselves to investigate how the horizontal advection terms not addressed in this study could be explicitly resolved.

ERF-VCV is applicable to retrieve volume-controlled NSAE across a target area. This practically permits rectifying footprint bias for eddy covariance flux in model data comparison, and provides the potential direction to improve energy budget closure in eddy covariance technique.

3.6 Acknowledgements

This work was supported by NEON, Inc. contract #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. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by Battelle Ecology, Inc. This material is based upon work supported by the National Science Foundation [grant DBI-0752017]. Any

opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

3.7 References

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. *Bulletin of the American Meteorological Society, 82*(11), 2415-2434. doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., . . . Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research: Biogeosciences (2005–2012), 116*(G2). doi:10.1029/2010jg001593

Brock, F. V. (1986). A nonlinear filter to remove impulse noise from meteorological data. *Journal of Atmospheric and Oceanic Technology, 3*(1), 51-58. doi:10.1175/1520-0426(1986)003<0051:ANFTRI>2.0.CO;2

Cacuci, D. G. (2003). *Sensitivity & Uncertainty Analysis, Volume 1: Theory*: CRC Press doi:10.1201/9780203498798.

Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-P., & Foken, T. (2014). Extension of the Averaging Time in Eddy-Covariance Measurements and Its Effect on the Energy Balance Closure. *Boundary-Layer Meteorology, 152*(3), 303-327. doi:10.1007/s10546-014-9922-6

Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., . . . Flanagan, L. B. (2011). Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling. *Agricultural and Forest Meteorology, 151*(1), 87-100. doi:10.1016/j.agrformet.2010.09.005

Croux, C., & Rousseeuw, P. J. (1992). *Time-efficient algorithms for two highly robust estimators of scale*: Springer doi:10.1007/978-3-662-26811-7_58.

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., . . . Stauch, V. J. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition

techniques. *Agricultural and Forest Meteorology, 148*(6-7), 821-838. doi:10.1016/j.agrformet.2007.11.012

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. *Agricultural and Forest Meteorology, 201*, 61-75. doi:10.1016/j.agrformet.2014.10.017

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. *Journal of Geophysical Research: Biogeosciences, 119*(3), 286-300. doi:10.1002/2013JG002392

Eder, F., De Roo, F., Rotenberg, E., Yakir, D., Schmid, H. P., & Mauder, M. (2015). Secondary circulations at a solitary forest surrounded by semi-arid shrubland and their impact on eddy-covariance measurements. *Agricultural and Forest Meteorology, 211–212*, 115-127. doi:http://dx.doi.org/10.1016/j.agrformet.2015.06.001

Finnigan, J. (2004). A re-evaluation of long-term flux measurement techniques part II: coordinate systems. *Boundary-Layer Meteorology, 113*(1), 1-41. doi:10.1023/B:BOUN.0000037348.64252.45

Finnigan, J. (2008). An introduction to flux measurements in difficult conditions. *Ecological Applications, 18*(6), 1340-1350. doi:10.1890/07-2105.1

Finnigan, J. J., Clement, R., Malhi, Y., Leuning, R., & Cleugh, H. A. (2003). A re-evaluation of long-term flux measurement techniques - Part I: Averaging and coordinate rotation. *Boundary-Layer Meteorology, 107*(1), 1-48. doi:10.1023/a:1021554900225

Foken, T. (2008). The energy balance closure problem: an overview. *Ecological Applications, 18*(6), 1351-1367. doi:10.1890/06-0922.1

Foken, T., Aubinet, M., Finnigan, J. J., Leclerc, M. Y., Mauder, M., & Paw U, K. T. (2011). Results of a panel discussion about the energy balance closure correction for trace gases. *Bulletin of the American Meteorological Society, 92*(4), ES13-ES18. doi:10.1175/2011BAMS3130.1

Foken, T., & Nappo, C. J. (2008). *Micrometeorology*: Springer

Foken, T., Wimmer, F., Mauder, M., Thomas, C., & Liebethal, C. (2006). Some aspects of the energy balance closure problem. *Atmospheric Chemistry and Physics, 6*, 4395-4402.

Homer, C., Dewitz, J., Yang, L. M., Jin, S., Danielson, P., Xian, G., . . . Megown, K. (2015). Completion of the 2011 National Land Cover Database for the Conterminous United States - Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering and Remote Sensing, 81*(5), 345-354. doi:10.14358/pers.81.5.345

Kanda, M., Inagaki, A., Letzel, M. O., Raasch, S., & Watanabe, T. (2004). LES study of the energy imbalance problem with Eddy covariance fluxes. *Boundary-Layer Meteorology, 110*(3), 381-404. doi:10.1023/B:BOUN.0000007225.45548.7a

Leuning, R., van Gorsel, E., Massman, W. J., & Isaac, P. R. (2012). Reflections on the surface energy imbalance problem. *Agricultural and Forest Meteorology, 156*, 65-74. doi:10.1016/j.agrformet.2011.12.002

Mahrt, L. (1998). Flux sampling errors for aircraft and towers. *Journal of Atmospheric and Oceanic Technology, 15*(2), 416-429. doi:10.1175/1520-0426(1998)015<0416:fsefaa>2.0.co;2

Mauder, M., Desjardins, R. L., Pattey, E., Gao, Z., & van Haarlem, R. (2008). Measurement of the Sensible Eddy Heat Flux Based on Spatial Averaging of Continuous Ground-Based Observations. *Boundary-Layer Meteorology, 128*(1), 151-172. doi:10.1007/s10546-008-9279-9

Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., . . . Desai, A. R. (2017). eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5. *Geoscientific Model Development, 10*(9), 3189-3206. doi:10.5194/gmd-10-3189-2017

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. doi: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. doi:10.5194/bg-10-2193-2013

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 Biology, 18*(2), 566-584. doi:10.1111/j.1365-2486.2011.02562.x

Rousseeuw, P. J., & Verboven, S. (2002). Robust estimation in very small samples. *Computational Statistics & Data Analysis, 40*(4), 741-758. doi:10.1016/S0167-9473(02)00078-6

Sakai, R. K., Fitzjarrald, D. R., & Moore, K. E. (2001). Importance of low-frequency contributions to eddy fluxes observed over rough surfaces. *Journal of Applied Meteorology, 40*(12), 2178-2192. doi:10.1175/1520-0450(2001)040<2178:iolfct>2.0.co;2

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. *Journal of Geophysical Research: Biogeosciences, 117*(G3). doi:10.1029/2012jg001960

Schlegel, F., Stiller, J., Bienert, A., Maas, H.-G., Queck, R., & Bernhofer, C. (2014). Large-Eddy Simulation Study of the Effects on Flow of a Heterogeneous Forest at Sub-Tree Resolution. *Boundary-Layer Meteorology, 154*(1), 27-56. doi:10.1007/s10546-014-9962-y

Schmid, H. P. (1997). Experimental design for flux measurements: matching scales of observations and fluxes. *Agricultural and Forest Meteorology, 87*(2-3), 179-200. doi:10.1016/s0168-1923(97)00011-7

Schotanus, P., Nieuwstadt, F., & De Bruin, H. (1983). Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes. *Boundary-Layer Meteorology, 26*(1), 81-93. doi:10.1007/BF00164332

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. *Journal of Geophysical Research: Atmospheres, 120*(4), 1348-1360. doi:10.1002/2014JD022338

Steinfeld, G., Letzel, M. O., Raasch, S., Kanda, M., & Inagaki, A. (2007). Spatial representativeness of single tower measurements and the imbalance problem with eddy-covariance fluxes: results of a large-eddy simulation study. *Boundary-Layer Meteorology, 123*(1), 77-98.

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. *Agricultural and Forest Meteorology, 171*, 137-152. doi:10.1016/j.agrformet.2012.11.004

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society, 79*(1), 61-78. doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Wilczak, J. M., Oncley, S. P., & Stage, S. A. (2001). Sonic anemometer tilt correction algorithms. *Boundary-Layer Meteorology, 99*(1), 127-150. doi:10.1023/A:1018966204465

Xu, K., Metzger, S., & Desai, A. R. (2017). Upscaling tower-observed turbulent exchange at fine spatio-temporalresolution using environmental response functions. *Agricultural and Forest Meteorology, 232*, 10-22. doi:10.1016/j.agrformet.2016.07.019

3.8 Tables

Table 3 Data filtering steps and the remaining percentage of data after each filtering step.

Processing step	Data filtering standard	Data remaining %
De-spiking	median filter de-spiking after Brock (1986)	99.5%
Data quality	3 hour periods with missing data >20% were discarded	60%
Unreliable nighttime data removal	Measurements at 122 m were discarded when sensible heat measured at 30 m < -10 W/m^2	50.4%
Spike removal	Lowest and highest 1% of flux values were removed	49.9%
Integral Turbulence Characteristics (ITC) test	Vertical velocity ITC <250% were selected	49.8%

Table 4 Comparison between aggregated tower observed turbulent flux and volume-controlled NSAE over the study period. Numbers in parenthesis are standard deviation of the diurnal cycle.

	H (Wm^{-2})	LE (Wm^{-2})	H + LE (Wm^{-2})
Tower observed turbulent flux	24.1(±149.5)	64.4(±256.6)	88.5
Volume-controlled NSAE	32.3(±120.9)	74.4(±228.9)	106.7

3.9 Figures

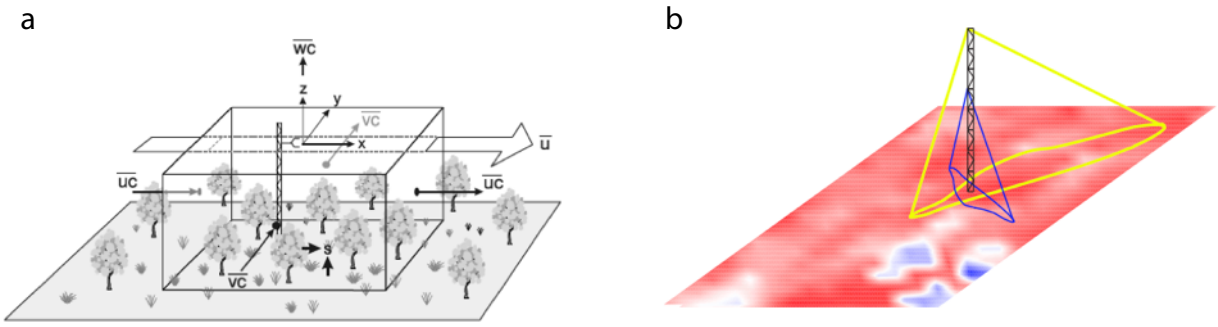


Figure 15 Conceptual plot of eddy covariance measurements in ideal condition (a) after Finnigan (2004) and in reality (b). Ideally, tower measurements can represent the net surface atmosphere exchange across the virtual control volume around it (a). However, in reality tower measurements can only represent a cone-shape that is confined by the footprint area over a heterogeneous surface.

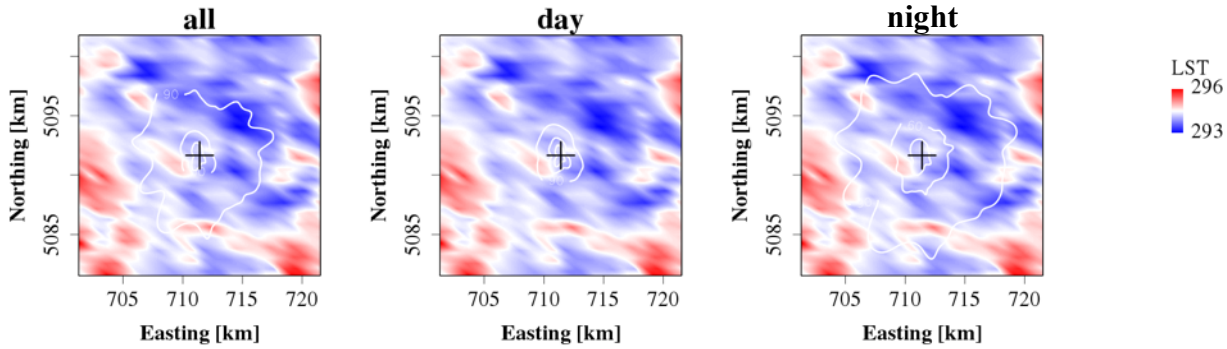


Figure 16 Footprint climatology (30%, 60% and 90%, white contour lines) for 122 m level measurements superimposed over average MODIS land surface temperature (LST) within 20×20 km^2 target domain surrounding the tower, which is indicated with the central crosshairs. (a) entire study time period; (b) daytime (9:00 – 17:00 CST), and (c) nighttime (17:00 – 9:00 CST). In this study, the footprint had a diurnal cycle and LST over the footprint area was lower than over the whole domain.

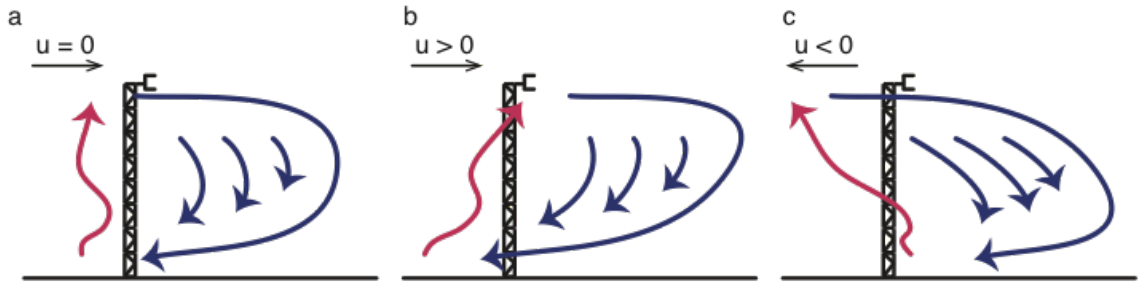


Figure 17 Conceptual figure of a tower measuring (a) no mesoscale circulation, (b) the updraft branch and (c) the downdraft branch of mesoscale circulation. Red arrows are updrafts and blue arrows are downdrafts.

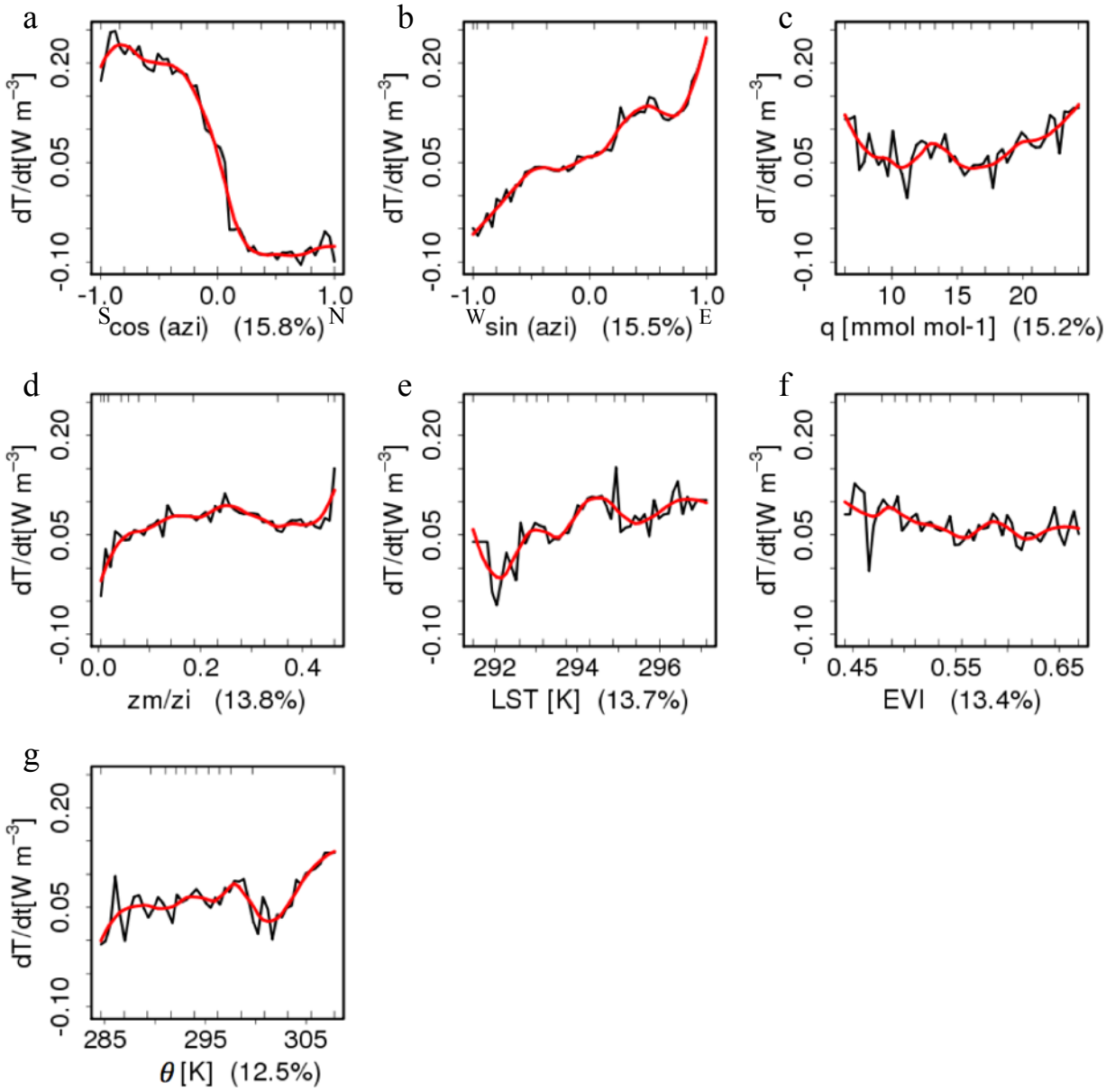


Figure 18 Equidistant response plots of temperature (T) time-rate of change, from which heat storage is derived. In decreasing order of explained variation in the flux signal (partial R^2 in braces), the drivers are cosine of the azimuth angle (cos(azi)), sine of the azimuth angle (sin(azi)), water vapor mixing ratio (q), relative height within boundary layer (z_m/z_i), land surface temperature (LST), enhanced vegetation indix (EVI), and potential temperature (θ) . 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 in red bold lines. The equidistant response plots use uniformly distributed percentiles within the range of training data (inward tickmarks on the lower x-axis). Inward tickmarks on the upper abscissa represent training data percentiles. (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.

.

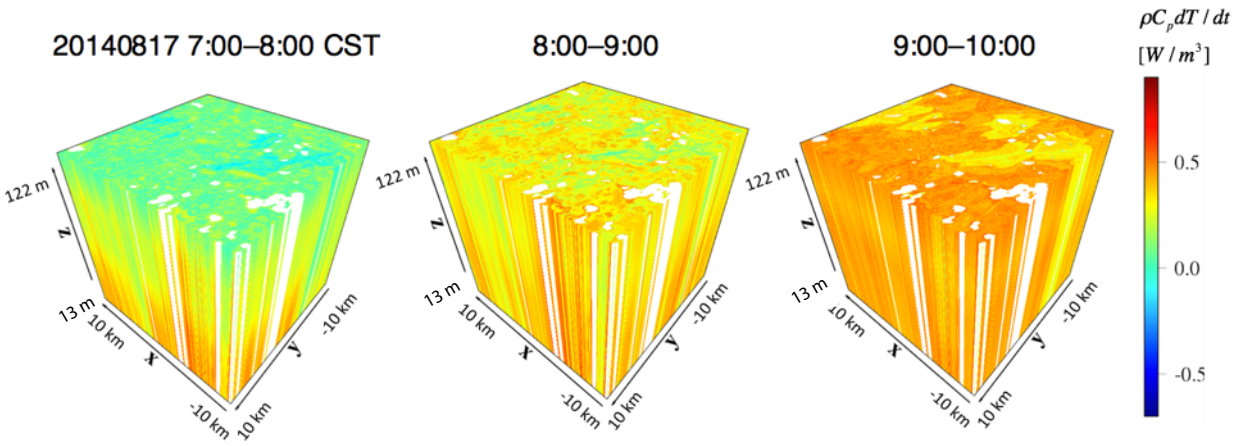


Figure 19 Volume projection of heat storage flux from the displacement height to the 122 m measurement height, across the tower-centered 20×20 km^2 target region. The color is the temperature time-rate-of-change in Wm^{-3}. The volume projections show a developing convective boundary layer from 2014 Aug 17 7:00 a.m.–10:00 p.m. CST. White spaces are time-space locations that cannot be projected as they exceed the range of the training data.

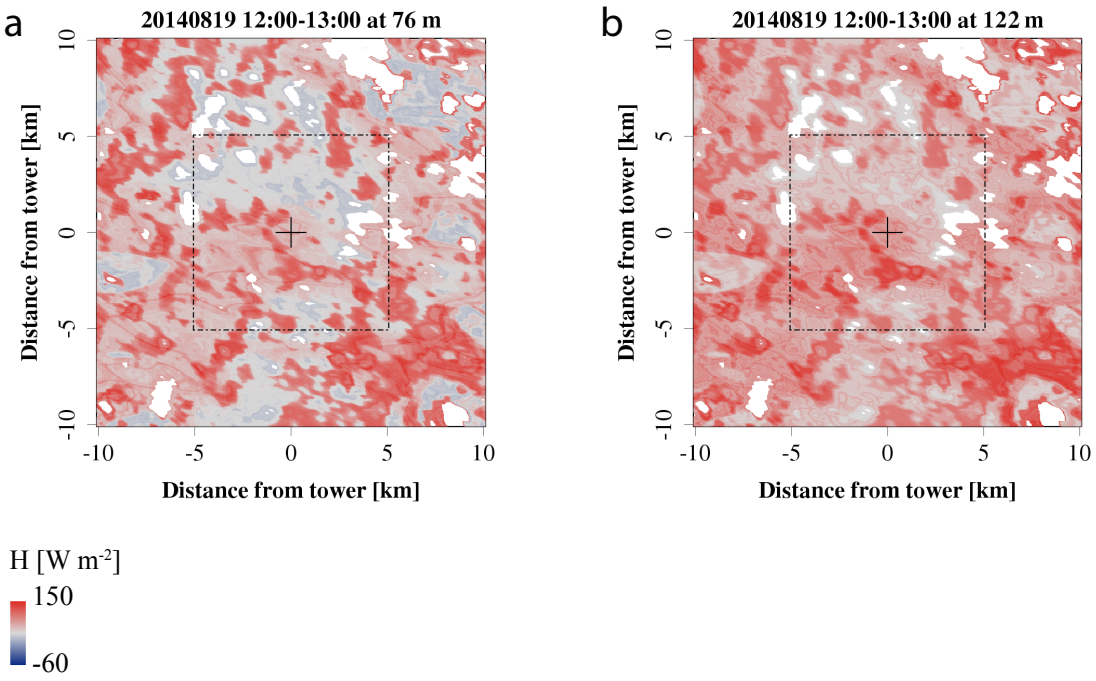


Figure 20 Projected vertical flux grids of sensible heat for August 19th, 2014, 12:00-13:00 CST for the tower-centered 20×20 km^2 target region at (a) 76 m and (b) 122 m level. Subplot (a) may indicate suitable conditions for surface heterogeneity-induced mesoscale circulations. Subplot (b) shows the flux grids at a higher elevation with less spatially distinct features due to blending.

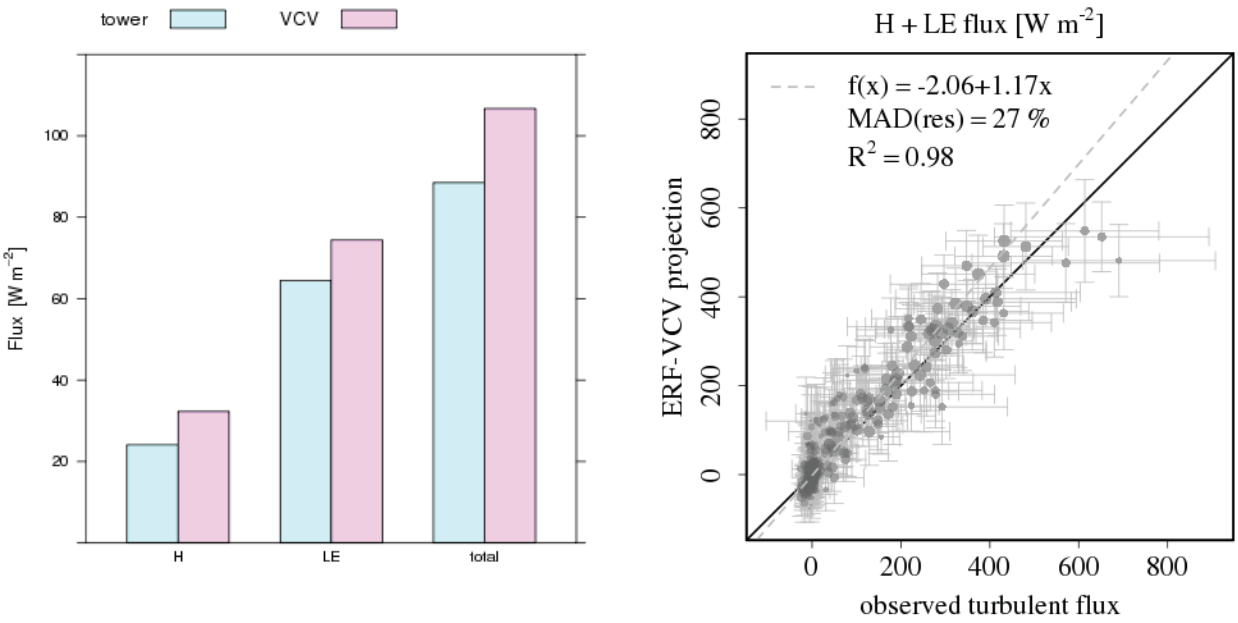


Figure 21 a. Tower observed turbulent flux (light blue) and VCV-estimated net surface atmosphere exchange (violet) for H, LE and projected net surface atmosphere exchanges (NSAE) integrated over the study period. ERF-VCV estimated NSAE sum of sensible and latent heat flux is 20.6% greater than tower observation when aggregated over the study period. b. Scatterplot for the observed vertical turbulent heat flux measurements for the sum of H and LE and ERF-VCV projected heat flux. Each point represents a one hour averaging period. Uncertainties are one standard deviation of the random uncertainties (error bars in x and y direction). Due to spatial averaging, ERF-VCV projections are less scattered. The fitted linear relationship shows that the ERF-VCV projections have a larger daytime-nighttime amplitude.

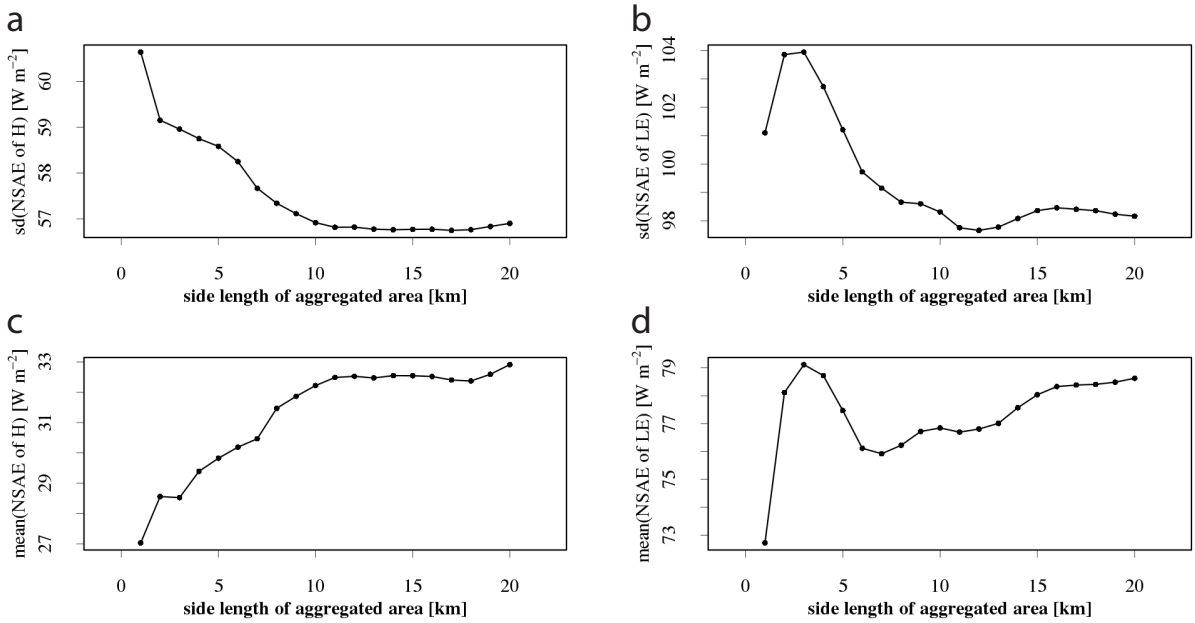


Figure 22: Temporal standard deviation (top) and mean (bottom) of hourly area-aggregated net surface flux as a function of the control volume side length. The left and right panels show net sensible and latent heat fluxes, respectively. Temporal standard deviation decreased and stabilized at spatial transporting scale of 10–15 km.

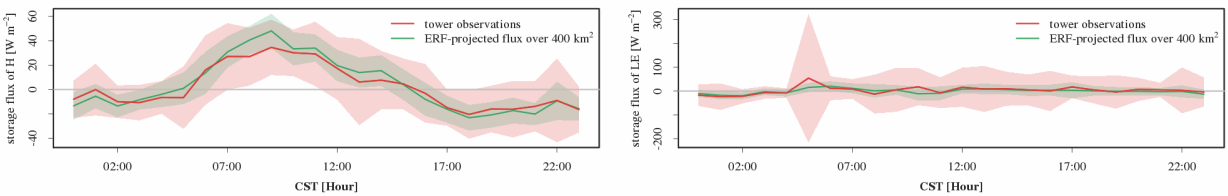


Figure 23: diurnal cycle of ERF-VCV projected storage flux (green) and tower observed storage flux (red) for H and LE generated from all data for the study period.

4 Can data mining help eddy-covariance see the landscape? A large-eddy simulation study[3]

4.1 Introduction

The eddy-covariance technique, in theory, can provide reliable, direct, spatially distributed and temporally continuous observations for surface-atmosphere exchanges of carbon, water and energy across contrasting eco-climate regions (Metzger, 2017). Now EC observations, such as those collected by AmeriFlux and National Ecological Observatory Network (NEON), have become available at unprecedented temporal duration and distributed spatial extents (Novick, et al., 2018). Near continuous (hourly/half-hourly) data on exchanges of carbon, water, heat and momentum are collected across the globe. The longest running towers are now approaching three decades of observations (Baldocchi, 2008). This drives EC to be one of the most important datasets for benchmarking ESMs (Bonan et al., 2011). Recent large land surface model intercomparisons as part of the North American Carbon Program diagnosed limitations in models, including spring phenology (Richardson et al., 2012), light use efficiency (Schaefer et al., 2012), and drought sensitivity (Schwalm et al., 2010). Wavelet coherence spectral analyses suggest that models have consistent biases at the diurnal and interannual timescale, for which eddy-covariance towers are well suited observations to validate and improve the simulation of carbon cycle in ESMs (Dietze et al., 2011; Stoy et al., 2013).

However, two challenges limit how well EC can reliably inform model-data comparison, assimilation, and data-inter-comparison and validation. The first one stems from footprint bias. This surface influence sampling footprint from EC towers not only varies rapidly in time, but

[3]This chapter is modified and excerpted with permission from *Xu, K., Sühring, M., Metzger, S., Durden, D., & Desai, A. R. Can data mining help eddy-covariance see the landscape? A large-eddy simulation study, In Preparation ready for submission.*

the spatial scale of flux tower footprint (10^{-1}–10^{1} km^2) is typically a mismatch with either the resolution of most earth system models (10^{2}–10^{4} km^2) or many remote sensing data product spatial resolutions (10^{-5}–10^{2} km^2), complicating the interpretation in the comparison with RS data and model-data comparison. Previous studies also show towers tend to under-sample warm convective spots, but over-sample cold subsidence areas, which is another source in location bias (Foken 2008; Wyngaard et al., 1984, Moeng et al., 1984).

The second challenge originates from longstanding challenge of energy balance non-closure in EC, incurred by unaccounted mesoscale eddies or under-sampling of hot spots by flux towers. Energy balance closure refers to a long-standing observed pattern that, at almost all sites, the observed turbulent sensible and latent heat fluxes are always 10–30% less than the sum of incoming available energy (net shortwave and long wave radiation minus ground heating), when expected to be zero at long term scale, e.g. multiple-years (Foken *et al*., 2011). One main reason is that low-frequency mesoscale flux contributions are hidden in the advection term, and inherently not captured by eddy-covariance technique as confirmed in a growing body of literature (e.g. Finnigan, 2003; Kanda et al., 2004; Foken 2008; Eder et al., 2015).

Atmospheric structures, i.e., turbulent organized structures (Kanda et al., 2004; Finnigan 2003) and/or secondary circulation associated with surface heterogeneity (Schlegel et al., 2014; Eder et al., 2015), can cause low-frequency mesoscale flux contributions. Turbulent organized structures are thermal-driven open cells (Eder et al, 2015; Wilczak and Tillman, 1980) or, in case of background wind, horizontal rolls (Drobinski et al., 1998; Maronga and Raasch, 2013) and typically have timescales larger than the averaging time of the eddy-covariance (Sakai et al., 2001; Foken et al., 2006). The under-sampling of hot spots by flux towers can also contribute to the energy imbalance (Foken, 2008; Wyngaard et al., 1984).

On occasionally but not always measured term, the atmospheric storage flux (Eqn. 1, term 1), does have a diurnal cycle and influences energy closure at hourly or finer temporal scale, despite its small averaged magnitude over daily timescale (Leuning *et al*., 2012). Especially in the case of tall towers or tall vegetated canopies, storage beneath the turbulent flux measurement height can comprise a substantial amount of the actual surface-atmosphere exchange at hourly resolution. Moreover, the flux footprint of turbulent flux measured at the tower top is mismatched with footprint of storage flux measured below, which is a potential source of flux error, too (Schmid 1997).

Previous studies found that adding more towers (Steinfeld et al., 2007; Mauder et al., 2008) or applying advanced scaling methods (ERF, Metzger, 2013, 2017; Xu et al., 2017a,b) can include mesoscale eddies in eddy-covariance measurements to help close energy budget (Kanda et al., 2004; Eder et al., 2015). However, these studies have not systematically evaluated advanced approaches for upscaling, rectifying, or combining multiple flux towers. Here, we attempt a first-of-its kind "Virtual flux tower" investigation of resolving spatial mismatch and energy balance closure using advanced upscaling techniques, and ask:

- How many eddy flux towers are needed to sufficiently rectify location bias, close energy budget, and sample the regional domain?
- Can an advanced scaling approach reduce this observational requirement, while still adequately sampling the regional flux domain?
- How accurate can advanced upscaling methods retrieve the spatial pattern of surface fluxes?

4.2 Methodology

In order to investigate the three questions raised in Sect. 4, we applied three different methods (Sect. 4.2.2), spatial EC, spatial-temporal EC, and ERF, to 384 virtual towers in Large-eddy Simulation (LES) (Sect. 4.2.1) with mesoscale eddies generated from prescribed stripe-like heterogeneous surface forcing.

4.2.1 LES set up and virtual tower measurements

Previous studies using 'real' measurements have difficulty detecting and distinguishing the source of systematic errors in tower measurements with unknown boundary conditions (Sühring et al., 2018; Schröter et al., 2000; Sühring and Raasch, 2013). With heterogeneous surface forcing in LES set up, this study targets to reveal the source and different systematic errors and magnitude of them to improve measurement strategies.

Here, the parallelized LES Model PALM (Raasch and Schroöter, 2001; Maronga et al., 2015; Suhring et al., 2018) was used for the numerical simulations in this study. The surface forcing was set to be stripe-wise alternating patches of higher sensible heat H (lower latent heat) and lower H (higher LE) along the x-axis with patch sizes L of 3 km., as illustrated in Figure 24a, b; hereafter refer to as warm-dry and cold-wet patches, respectively. Along the y-axis, the model surface was homogeneous. The domain-averaged surface heat fluxes were constant in time with a value of H=0.1 Kms^{-1} and LE=6 x 10^{-5} kg $kg^{-1}ms^{-1}$ (each equals a heat flux of $100Wm^{-2}$, assuming a constant air density of 1 x 10^{3} kgm^{-3} in all performed simulations. Stripe-like surface patches with the surface sensible heat alternated between 150 and 50 W m-2 and latent heat alternated between 150 and 250 W m-2 (Sühring et al., 2014). With horizontal wind perpendicular to the surface heterogeneity, secondary circulations can develop (Figure 24c).

PALM LES was set up over a domain of 12 km x 16 km x 1.8 km at 7 m spatial resolution and at a time step of 0.3 s. The simulations ran for 5 hours of simulation time, while data analysis started after 2 hours. Detailed LES set up and boundary layer structure can be found in Suhring et al., (2018). 384 virtual flux towers (w,q,t,p) were sampled at 49 m vertical level (Figure 24, Suring et al., 2018).

4.2.2 Analysis

We applied four different methods, spatial EC, spatial-temporal EC (S07 and M08), and ERF, to reproduce the domain flux mean and variation.

4.2.2.1 Spatial eddy-covariance

Traditional eddy-covariance using stationary towers is calculated in temporal domain with only one tower. With multiple towers set up in this study, eddy-covariance representative flux can be calculated in spatial domain (Equation 1).

$$\overline{[F_{spatial\ EC}]} = \overline{[w_s'(x,y,t) \cdot c_s'(x,y,t)]} = \overline{[(w(x,y,t) - [w(x,y,t)]) \cdot (c(x,y,t) - [c(x,y,t)])]} \quad \text{Equation 1}$$

Here, *c* represents air temperature, T and water vapor mixing ratio, q. [ϕ] denotes the horizontal average of a quantity ϕ over a certain horizontal domain and if ϕ denotes the temporal average of a quantity ϕ over a certain period of time. According to Reynolds decomposition (Foken and Nappo, 2008, Stull, 1988), w and c can be split up into a horizontal mean and a related deviation (marked by a prime and subscript s), i.e. $w_s'(x,y,t)$ and $c_s'(x,y,t)$.

After subtracting, aggregate $w_s'(x,y,t)$ and $c_s'(x,y,t)$ over half an hour for each tower, then aggregate over all towers in the horizontal domain. This calculation was repeated 50,000 times to obtain ensemble representative flux.

4.2.2.2 Spatio-temporal eddy-covariance

We derived spatial-temporal eddy-covariance method from Steinfeld et al., (2007, S07).

From Equation 23 in Steinfeld et al., (2007), representative flux is calculated as:

$$\overline{[F_{S07}]} = \overline{[w \cdot ([c] - [\bar{c}])]} + \overline{[w \cdot (c - [c])]} = \overline{[wc]} - \overline{[w][c]} \qquad \text{Equation 2}$$

Note that if we calculate flux using spatio-temporal eddy covariance, we get the same result. Following is the derivation.

$$\overline{[\boldsymbol{F_{spatio-\square emporal\ EC}}]} = \overline{[\boldsymbol{w_{st}}'(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}) \cdot \boldsymbol{c_{st}}'(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t})]} \qquad \text{Equation 3}$$

while,

$$w_{st}'(x,y,t) = w(x,y,t) - \overline{[w]} \qquad \text{Equation 4}$$

$$c_{st}'(x,y,t) = c(x,y,t) - \overline{[c]} \qquad \text{Equation 5}$$

when using Reynold decomposition,

$$\overline{[\boldsymbol{F_{spatio-temporal\ EC}}]} = \overline{\left[\left(\boldsymbol{w}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}) - \overline{[\boldsymbol{w}]}\right) \cdot \left(\boldsymbol{c}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}) - \overline{[\boldsymbol{c}]}\right)\right]} \qquad \text{Equation 6}$$

After expanding Equation 6

$$\overline{[\boldsymbol{F_{spatio-temporal\ EC}}]} = \overline{[\boldsymbol{wc} - [\bar{\boldsymbol{w}}]\boldsymbol{c} - \boldsymbol{w}[\bar{\boldsymbol{c}}] + [\bar{\boldsymbol{w}}][\bar{\boldsymbol{c}}]]} = [\overline{\boldsymbol{wc}}] - [\bar{\boldsymbol{w}}][\bar{\boldsymbol{c}}] - [\bar{\boldsymbol{w}}][\bar{\boldsymbol{c}}] + [\bar{\boldsymbol{w}}][\bar{\boldsymbol{c}}] = [\overline{\boldsymbol{wc}}] - [\bar{\boldsymbol{w}}][\bar{\boldsymbol{c}}] = \overline{[\boldsymbol{F_{S07}}]} \qquad \text{Equation 7}$$

As S07 and spatio-temporal EC are identical, here we call S07 method as “spatio-temporal EC”. Mauder et al., 2008’s method built on S07, but with only one high-frequency of the tower, for rest, temperature and humidity were measured at low-frequency. Therefore, here we call S07

and M08 methods by a joint name “spatio-temporal EC”. The averaging time period is half-hour for S07 and M08. The ensemble size is 50,000.

4.2.2.3 Environmental Response Function (ERF)

The underlying principle of ERF (Metzger et al., 2013; Xu et al., 2017a) 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.

In short, the essential steps of ERF are: (i) observing key environmental drivers and responses at high spatiotemporal resolution; (ii) inferring process relationships among environmental drivers and responses using artificial intelligence; (iii) projecting environmental responses in 3-D space and time.

In ERF-VCV, these three essential steps are realized by coupling time-frequency-decomposed eddy-covariance flux observations (Xu et al., 2015) with flux footprint modelling (Kljun et al., 2015) and machine learning (Elith et al., 2008). The traditional eddy-covariance approach requires averaging turbulent fluctuations over a 30 to 60-minute window, making it susceptible to changing turbulent conditions and footprint variations that occur on shorter time scales. By using a wavelet-based spectral average instead, the flux measurement period can be reduced substantially (5-minute window size at 1-minute resolution), permitting clear spatial attribution without neglecting long-wavelength flux contributions. The resulting large sample size and high signal-to-noise ratio enables machine learning to extract key relationships between these atmospheric flux ‘responses’, and land surface and meteorological ‘drivers’.

In this study, ERF was applied to randomly-chosen 1 to 14 towers for 2000 ensemble times. We used time series of hour 3–4, with transporting scale in the wavelet decomposition is 1 hour, in order to get rid of edge-effect in wavelet decomposition. With n virtual tower(s), whereas n is from 1 to 14, the training dataset for machine learning is 60×n.

Drivers consisted of surface properties, meteorological variables, and relative height within the boundary layer (z_m/z_i). Meteorological variables were air temperature, water vapor mixing ratio measured by virtual tower. Surface properties were land surface temperature (LST) and land surface moisture (LSM) at 3.5 m derived from x-z cross sections, the lowest level above the surface. The two sets of temperature and humidity drivers were used to indicate the gradient between the air and surface.

Storage flux were calculated as the domain averaged time-rate of change of air temperature and humidity using x-z cross sections.

4.3 Results and Discussion

4.3.1 How many flux towers are needed to sufficiently sample the flux domain mean?

When using spatial EC, adding each tower helps ensemble averaged H close the energy budget (orange line in Figure 26a). But a relatively high tower density is needed to fully close the energy: when 14 towers are used per 12 km×16 km (7.3 towers per 100km^2), the represented flux only reach 80% of the applied surface flux. After account for 5.6% of energy transformed to storage flux, the spatial-EC method is still unable to capture and explain 13.5% of H + LE flux. Spatio-temporal EC (S07 and M08 in Figure 26a) performs much better: with one single tower, ensemble averaged H starts from 79.2%, also increases with more towers, and reaches 85% of energy with fewer (5) towers, leaving 9.2% of total surface energy balance flux unexplained.

Similar results are seen in LE (Figure 26b, note the values at y axis in Figure 26b should be divided by the surface forcing of LE, 200 Wm^{-2}, to get relative percentage): ensemble average LE derived from spatio-temporal EC increases from 91.1% to 94.6% of the expected LE surface flux as one goes from X to Y towers, indicating the contribution from mesoscale eddies in LE limit the use of single towers. Spatio-temporal EC performs better than spatial EC in terms of the represented flux with one single tower and the numbers of towers required to close the total LE flux.

Spatial variation of fluxes could be derived from S07 and spatial EC in theory, if number of towers is larger than 2. But simple bi-linear interpolation into the whole domain would not give a good retrieval of the surface especially with towers < 20, because the surface patches are stripe-like. The companion paper (Suhring et al., 2018) used the same LES set up and flux disaggregation method using virtual airborne EC measurements to retrieve the spatial variation, and tested the retrievals with the surface forcing reference (finding what?)

Based on this analysis, we surmise that the detected mesoscale eddies contribute 6% and 3.5% of energy budget for H and LE, separately (**Table** 5). Also, storage flux in this convective situation is not negligible, 5.6% and 3.4% for H and LE (**Table** 5). When considering storage flux and mesoscale eddies in sum, spatio-temporal EC helps to close the energy budget from 20.8% to 9.2% for H, and from 8.9% to 2% for LE, with only 5 towers.

When compared with previous work, 5 towers are needed per 12 km×16 km region in this study is consistent with the findings in Steinfeld et al., (2007) and Mauder et al., (2008). We also proved the M08 method (Mauder et al., 2008) generates almost identical results with S07, but with only one high-frequency tower and remaining towers are measuring only mean quantities, thereby reducing the observation expense. In addition, while previous studies suggested that the evidence of latent heat-generated mesoscale eddies is lacking (e.g. Eder et al., 2015), this

LES experiment identified similar magnitude of mesoscale eddy contribution to H and LE underestimation. As discussed more in Sect. 4.3.2, evidence of secondary circulation generated by the horizontal gradient of H flux are shown, while the lack of evidence of LE-generated secondary circulation indicates that the mesoscale eddies in LE are generated as turbulent organized structures.

4.3.2 How does the upscaling method, ERF, help flux towers to sample the regional domain?

Upscaling and wavelet based flux computation significantly improves the reliability of energy flux measurement even with a single tower. With one single tower, the total ERF-projected flux of H and LE incorporates additional energy term of 4.3%, persurmably from mesoscale eddies, reducing the energy budget gap from 12.9% to 4.6% (Figure 26c, **Table** 5). This reduces the observation requirement by 80%, compared to five towers using spatio-temporal EC. As expected with the method, with more towers, ERF-projected ensemble total flux of H and LE remain unchanged (Figure 26c). This is achieved by the longer flux transporting scale enabled in the wavelet composition in ERF. Additionally, the ensemble standard deviation of ERF and spatio-temporal EC are about the same magnitude (figure not shown), reflecting the sensitivity of both methods to the choice of virtual towers is about the same.

If we look at ERF-projected H and LE separately, ERF can retrieve 85.2% of H and 93.9% of LE with one single tower. With 7 towers, ERF-projected ensemble energy increases to 92% for H, while decreases to 90.4% for LE (red line in Figure 26b). We found this decrease occurs because the warm-dry patches are under-sampled by virtual towers (green line in Figure 27a), and adding more towers in ERF increase the chance of warm dry patches to be detected and projected by ERF (yellow, orange, and red linesFigure 27a) while reducing the change of detecting cooler, wetter patches. Adding more towers especially improve the sampling of

warm-dry, spatially confined and strong convective patches, whereas spatially abundant cold-wet patches can be easily detected even with few towers. This finding is consistent with previous studies reflecting that in real world EC towers also tend to under-sample hot spots (Foken 2008; Wyngaard et al., 1984).

Under-sampling of hot spots by towers occurs as a result of atmospheric convection skewness. In the updraft branch of the mesoscale eddies, convective eddies gather more energy to overcome gravity, and become stronger and spatially-smaller as they go up, as observations show larger magnitude but less probability density (second peak of green line inFigure 27a) than the surface forcing reference (second peak of blue line Figure 27a). The weaker downward flow needs more area than the stronger upward flow so they compensate each other according to mass conservation (Figure 28). The corresponding spatial distribution of atmospheric fields is skewed towards weaker, cooler subsidence areas (as opposed to stronger, warmer convective areas), even though cold/warm surface forcing (real-world or virtual) are uniformly distributed in space (Figure 28). Because the EC tower is operating in this atmospheric field in a fixed temporal frame of reference, its measurement is inherently "biased" to capture more of the weaker downward flow as opposed to the stronger upward flow (Figure 28).

The atmospheric convection skewness also explains the decrease of ERF-projected LE with more towers (red line in Figure 26b). The tower-measured land surface moisture (LSM), 2.32 mkg/kg, is shown to be larger than domain mean LSM, 2.38 mkg/kg. But different from H, LE observations tend to be more mixed and homogenized with height? (green line in Figure 27b). This can be explained by the theory that heterogeneity-induced mesoscale eddies dominated by sensible heat fluxes mix the passive water vapor, making water vapor field more homogenized. An alternative explanation is that convergence of warm updraft branch (with low LE) and

divergence of cold downdraft branch (with high LE) lead to the increase the magnitude of lower LE, and decrease the magnitude of high LE, making tower-measured LE more homogenized.

ERF is able to rectify this location bias, while spatio-temporal EC cannot. The more towers ERF utilizes, the better it can rectify. Further, ERF not only increase the probability density of the warm-dry patches in H projection, it also enables reducing the magnitude of the converged warm fluxes due to atmospheric skewness (second peak of red line in Figure 27a). In this case, the location bias incurred by atmospheric skewness contribute to 5.8% and -3% energy for H and LE (**Table** 5).

The remaining 4.6% "missing" flux can be hidden in projected H, since the projection of H still cannot produce bimodal distribution (Figure 27a), which means warm-dry patches are fractionally under-projected.

4.3.3 How accurate can ERF retrieve the surface flux variation?

Based on the ERF calculated projection statistics, ERF can rectify location bias by improving flux coverage over the regional domain from <1% tower footprint area (based on footprint spatial coverage), to 10%, 42%, and 59%, respectively, with one, two and 14 virtual towers. When the surface properties exceed the range of what the tower(s) can project, ERF rejects to project to these areas (white areas in Figure 29), and this is how the coverage percentages are calculated. The low spatial coverage with one tower is likely arising from atmospheric condition being quasi-stationary in this LES set up, making each virtual tower measure have almost the same footprint area, an effect that is less likely in the real-world. Applying ERF over several months flux tower data would build an ensemble of different atmospheric conditions and thus a footprint climatology, which in turn improves the spatial representation, such as larger than 60% over 10×10 km^2 (Xu et al., 2017a).

The fitted relationship (dashed line in Figure 30) between ERF-projected H and known LES surface reference is very close to 1:1 line. These two show very good agreement with 13% of median absolute deviation of residual. The extracted response functions of H are physically reasonable (Figure 31). Both proved the theoretical soundness and statistical significance of ERF approach.

Because LE flux at 49 m virtual tower is much more homogenized than tower-measured H flux (Figure 25, Figure 27b), ERF is unable to distinguish the two different patch contributions to LE, gives less successful projection performance.

4.4 Conclusion

In our study, one and six virtual eddy-covariance towers, respectively, were able to capture only 79% and 91.5% of the half-hourly turbulent heat flux over a 12 km×16 km LES domain using the spatio-temporal EC method. Once the ERF upscaling method was applied, 91.5% and 98.3% of the domain mean sensible heat were captured with one single tower and with 7 towers, respectively. This is achieved because ERF resolves mesoscale eddies using wavelet-decomposed fluxes, which considers high-frequency (sub-hourly) fluxes with long transporting scales, e.g. one hour or three hours. We also find, in this convective LES set up, that storage flux contributes to 4.1% out of 13.2% eddy-covariance energy imbalance. Storage flux did not converge to zero as often assumed at diurnal cycle scale, especially during convection events such as in this LES experiment.

In this experiment, evidence does show mesoscale eddies in latent heat flux (LE). LE mesoscale eddies are not generated by LE itself. Rather, the updraft and downdraft branches generated by sensible heat fluxes trigger the mesoscale eddies in LE field.

Evidence of secondary circulation generated by the horizontal gradient of sensible heat flux (H) are shown, while the lack of evidence of latent heat (LE)-generated secondary circulation indicates that the mesoscale eddies in LE are generated as turbulent organized structures.

In addition, ERF rectifies the location bias introduced by atmospheric skewness: strong convective areas are sparsely distributed, while cool subsidence areas cover the majority of the space. But in this LES analysis, for the surface, warm and cold patches are evenly distributed. Convection and heterogeneity-induced mesoscale eddies tend to increase ratio of cold-wet patches to warm-dry patches observed by towers, driving towers to under-sample warm-dry area. ERF rectifies atmospheric-skewed location bias by retrieving more warm areas in mapping surface atmosphere exchanges, while traditional spatio-temporal eddy-covariance using more towers cannot either detect or correct this atmospheric-skewed location bias.

The theoretical soundness and statistical significance of ERF approach are proved by both the very good agreement between ERF-projection and known LES surface reference, and the physically reasonable extracted response functions. ERF enables rectifying location bias by improving flux coverage over the regional domain from <1% tower footprint area, to 20%, 62%, and 89%, respectively, with one, two and 14 virtual towers.

Future work can be on: 1) application of ERF with virtual towers in LES to resolve advection terms; 2) Comparison with other flux decomposition methods, e.g. flux disaggregation method, with LES models; 3) test ERF with flux tower cluster real observations; 4) test the hypothesis that with spatially more abundant updraft regions, the warm patches are not anymore under-sampled;

This paper demonstrates how advanced scaling techniques can captures effect of mesoscale eddies, maps fluxes across regional scale, and decrease the number of towers required for

sampling both mean and variation of heterogeneous surface, and thus decrease experimental expense.

4.5 Reference

Baldocchi, D. (2008). TURNER REVIEW No. 15.'Breathing'of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Australian Journal of Botany, 56(1), 1-26. doi:10.1071/BT07151

Bonan, G. B. (2008). Ecological climatology: concepts and applications. Cambridge, UK: Cambridge University Press

Dietze, M. C., Vargas, R., Richardson, A. D., Stoy, P. C., Barr, A. G., Anderson, R. S., . . . Weng, E. (2011). Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis. Journal of Geophysical Research: Biogeosciences, 116(G4), n/a-n/a. doi:10.1029/2011JG001661

Drobinski, P., Brown, R. A., Flamant, P. H., & Pelon, J. (1998). Evidence of organized large eddies by ground-based Doppler lidar, sonic anemometer and sodar. Boundary-Layer Meteorology, 88(3), 343-361. doi:10.1023/a:1001167212584

Eder, F., De Roo, F., Rotenberg, E., Yakir, D., Schmid, H. P., & Mauder, M. (2015). Secondary circulations at a solitary forest surrounded by semi-arid shrubland and their impact on eddy-covariance measurements. Agricultural and Forest Meteorology, 211–212, 115-127. doi:http://dx.doi.org/10.1016/j.agrformet.2015.06.001

Finnigan, J. J., Clement, R., Malhi, Y., Leuning, R., & Cleugh, H. A. (2003). A re-evaluation of long-term flux measurement techniques - Part I: Averaging and coordinate rotation. Boundary-Layer Meteorology, 107(1), 1-48. doi:10.1023/a:1021554900225

Foken, T. (2008). The energy balance closure problem: an overview. Ecological Applications, 18(6), 1351-1367. doi:10.1890/06-0922.1

Foken, T., Aubinet, M., Finnigan, J. J., Leclerc, M. Y., Mauder, M., & Paw U, K. T. (2011). Results of a panel discussion about the energy balance closure correction for trace gases.

Bulletin of the American Meteorological Society, 92(4), ES13-ES18. doi:10.1175/2011BAMS3130.1

Foken, T., & Nappo, C. J. (2008). Micrometeorology: Springer

Foken, T., Wimmer, F., Mauder, M., Thomas, C., & Liebethal, C. (2006). Some aspects of the energy balance closure problem. Atmospheric Chemistry and Physics, 6, 4395-4402.

Kanda, M., Inagaki, A., Letzel, M. O., Raasch, S., & Watanabe, T. (2004). LES study of the energy imbalance problem with Eddy covariance fluxes. Boundary-Layer Meteorology, 110(3), 381-404. doi:10.1023/B:BOUN.0000007225.45548.7a

Leuning, R., van Gorsel, E., Massman, W. J., & Isaac, P. R. (2012). Reflections on the surface energy imbalance problem. Agricultural and Forest Meteorology, 156, 65-74. doi:10.1016/j.agrformet.2011.12.002

Maronga, B., & Raasch, S. (2013). Large-Eddy Simulations of Surface Heterogeneity Effects on the Convective Boundary Layer During the LITFASS-2003 Experiment. Boundary-Layer Meteorology, 146(1), 17-44. doi:10.1007/s10546-012-9748-z

Mauder, M., Desjardins, R. L., Pattey, E., Gao, Z., & van Haarlem, R. (2008). Measurement of the Sensible Eddy Heat Flux Based on Spatial Averaging of Continuous Ground-Based Observations. Boundary-Layer Meteorology, 128(1), 151-172. doi:10.1007/s10546-008-9279-9

Metzger, S. (2017). Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations. Agricultural and Forest Meteorology(Massman special issue).

Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Widemann, B. T. Y., Neidl, F., . . . Foken, T. (2013). Spatially explicit regionalization of airborne flux measurements using environmental response functions. Biogeosciences, 10(4), 2193-2217. doi:10.5194/bg-10-2193-2013

Moeng, C.-H., & Wyngaard, J. C. (1984). Statistics of Conservative Scalars in the Convective Boundary Layer. Journal of the Atmospheric Sciences, 41(21), 3161-3169. doi:10.1175/1520-0469(1984)041<3161:socsit>2.0.co;2

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. Agricultural and Forest Meteorology, 249(Supplement C), 444-456. doi:https://doi.org/10.1016/j.agrformet.2017.10.009

Raasch, S., Schroter, M., Ams, & Ams. (2002). A large-eddy simulation model performing on massively parallel computers. Boston: Amer Meteorological Society

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 Biology, 18(2), 566-584. doi:10.1111/j.1365-2486.2011.02562.x

Sakai, R. K., Fitzjarrald, D. R., & Moore, K. E. (2001). Importance of low-frequency contributions to eddy fluxes observed over rough surfaces. Journal of Applied Meteorology, 40(12), 2178-2192. doi:10.1175/1520-0450(2001)040<2178:iolfct>2.0.co;2

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. Journal of Geophysical Research: Biogeosciences, 117(G3). doi:10.1029/2012jg001960

Schlegel, F., Stiller, J., Bienert, A., Maas, H.-G., Queck, R., & Bernhofer, C. (2014). Large-Eddy Simulation Study of the Effects on Flow of a Heterogeneous Forest at Sub-Tree Resolution. Boundary-Layer Meteorology, 154(1), 27-56. doi:10.1007/s10546-014-9962-y

Schmid, H. P. (1997). Experimental design for flux measurements: matching scales of observations and fluxes. Agricultural and Forest Meteorology, 87(2-3), 179-200. doi:10.1016/s0168-1923(97)00011-7

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. Journal of Geophysical Research: Biogeosciences, 115(G3). doi:10.1029/2009JG001229

Steinfeld, G., Letzel, M. O., Raasch, S., Kanda, M., & Inagaki, A. (2007). Spatial representativeness of single tower measurements and the imbalance problem with eddy-covariance fluxes: results of a large-eddy simulation study. Boundary-Layer Meteorology, 123(1), 77-98.

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. Agricultural and Forest Meteorology, 171, 137-152. doi:10.1016/j.agrformet.2012.11.004

Stull, R. B. (1988). An introduction to boundary layer meteorology (Vol. 13): Springer Science & Business Media

Sühring, M., Maronga, B., Herbort, F., & Raasch, S. (2014). On the Effect of Surface Heat-Flux Heterogeneities on the Mixed-Layer-Top Entrainment. Boundary-Layer Meteorology, 151(3), 531-556. doi:10.1007/s10546-014-9913-7

Sühring, M., & Raasch, S. (2013). Heterogeneity-Induced Heat-Flux Patterns in the Convective Boundary Layer: Can they be Detected from Observations and is There a Blending Height?—A Large-Eddy Simulation Study for the LITFASS-2003 Experiment. Boundary-Layer Meteorology, 148(2), 309-331. doi:10.1007/s10546-013-9822-1

Wilczak, J. M., & Tillman, J. E. (1980). THE 3-DIMENSIONAL STRUCTURE OF CONVECTION IN THE ATMOSPHERIC SURFACE-LAYER. Journal of the Atmospheric Sciences, 37(11), 2424-2443. doi:10.1175/1520-0469(1980)037<2424:ttdsoc>2.0.co;2

Wyngaard, J. C., & Brost, R. A. (1984). Top-Down and Bottom-Up Diffusion of a Scalar in the Convective Boundary Layer. Journal of the Atmospheric Sciences, 41(1), 102-112. doi:10.1175/1520-0469(1984)041<0102:tdabud>2.0.co;2

Xu, K., Metzger, S., & Desai, A. R. (2017a). Upscaling tower-observed turbulent exchange at fine spatio-temporalresolution using environmental response functions. Agricultural and Forest Meteorology, 232, 10-22. doi:10.1016/j.agrformet.2016.07.019

Xu, K., Metzger, S., & Desai, A. R. (2017b). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. Agricultural and Forest Meteorology. doi:10.1016/j.agrformet.2017.10.011

4.6 Table

Table 5 Energy budget for turbulent flux, ERF-resolved mesoscale eddies, storage flux, and the remaining part of sensible heat flux (H) and latent heat flux (LE).

	Turbulent flux	Mesoscale eddies	Storage flux	Atmospheric skewness	missing
H	79.2%	6%	5.6%	5.8%	3.4%
LE	91.1%	3.5%	3.4%	-3%	5%
H + LE	87.1%	4.3%	4.1%	-0.08%	4.6%

4.7 Figures

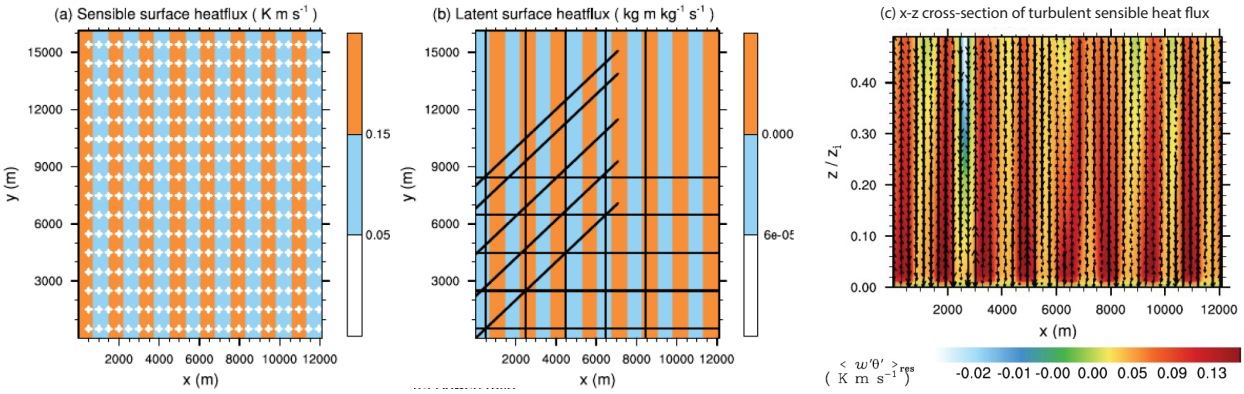


Figure 24 The heterogeneously prescribed surface-fluxes of a) sensible heat flux with white crosses indicating the horizontal position of the virtual tower measurements and b) latent heat flux with black lines indicating the horizontal flight tracks of virtual aircraft measurements (not used here). c) (x,z)-cross-section of the turbulent sensible heat flux (colored contours), as well as the corresponding flow field (vectors) at different points in time during the analysis period 3–4 h, indicates the mesoscale eddies occur with larger heat fluxes in the uprising branches over the warm-dry patches, and suppressed heat fluxes in the subsiding branches over the cold-wet patches.

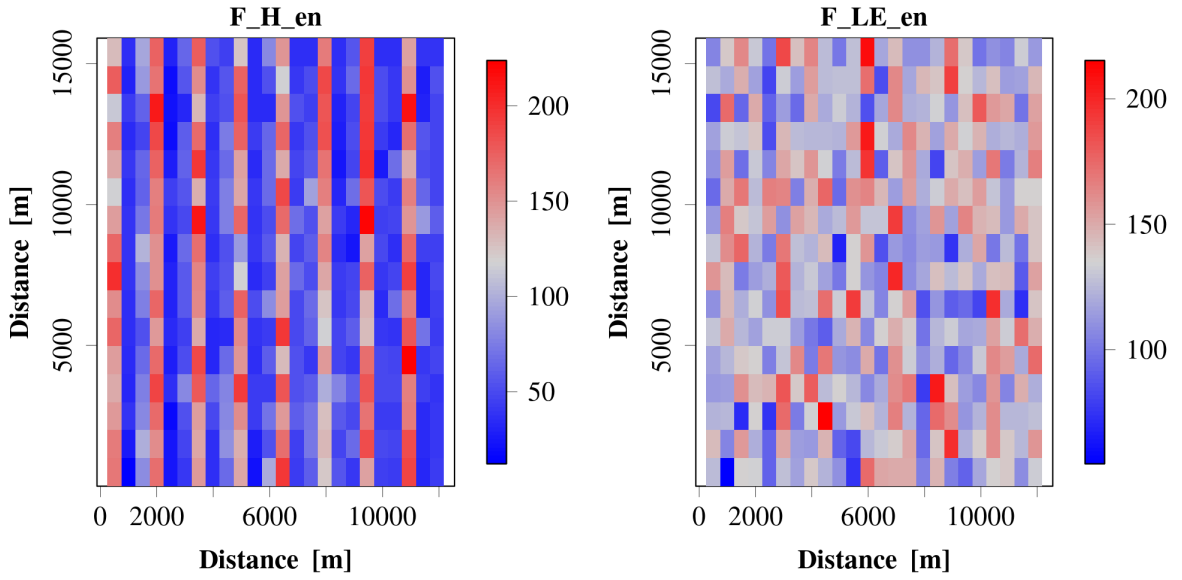


Figure 25: Virtual tower measured sensible heat and latent heat fluxes at 50 m. Sensible heat flux map reflects strong warm-cold patches, while latent heat flux maps are more mixed and homogenized without strong wet-dry contrast.

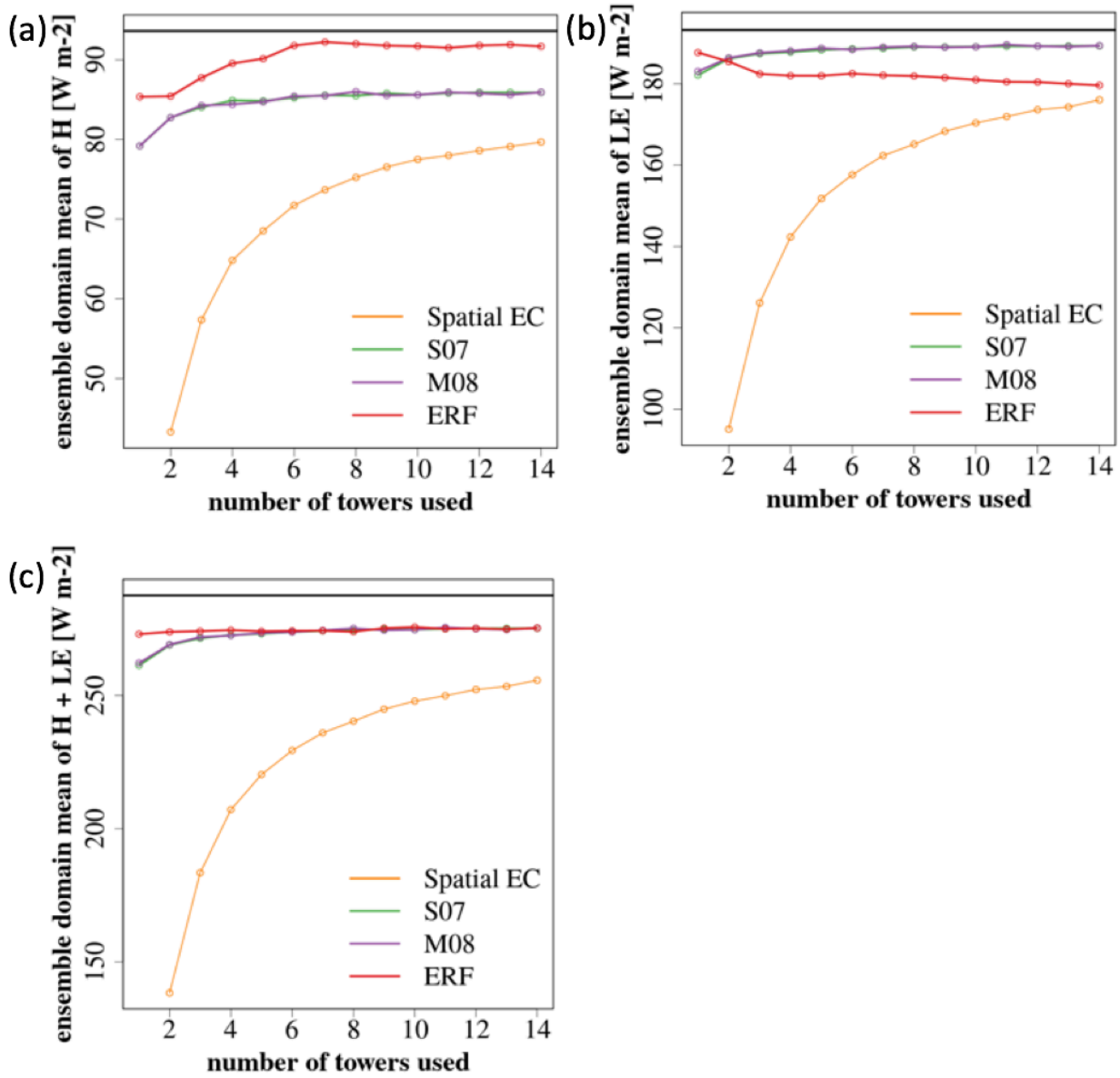


Figure 26 Ensemble averaged energy fluxes from randomly placed virtual towers at 49 m using spatio-temporal eddy-covariance method (S07 and M08) for a) sensible heat (H), b) latent heat (LE), and c) H + LE. ERF enables rectifying location bias for H and LE, while S07 and M08 cannot. ERF enable reaching 91.7% of H+LE with one single tower, while S07 and M08 needs 5 towers.

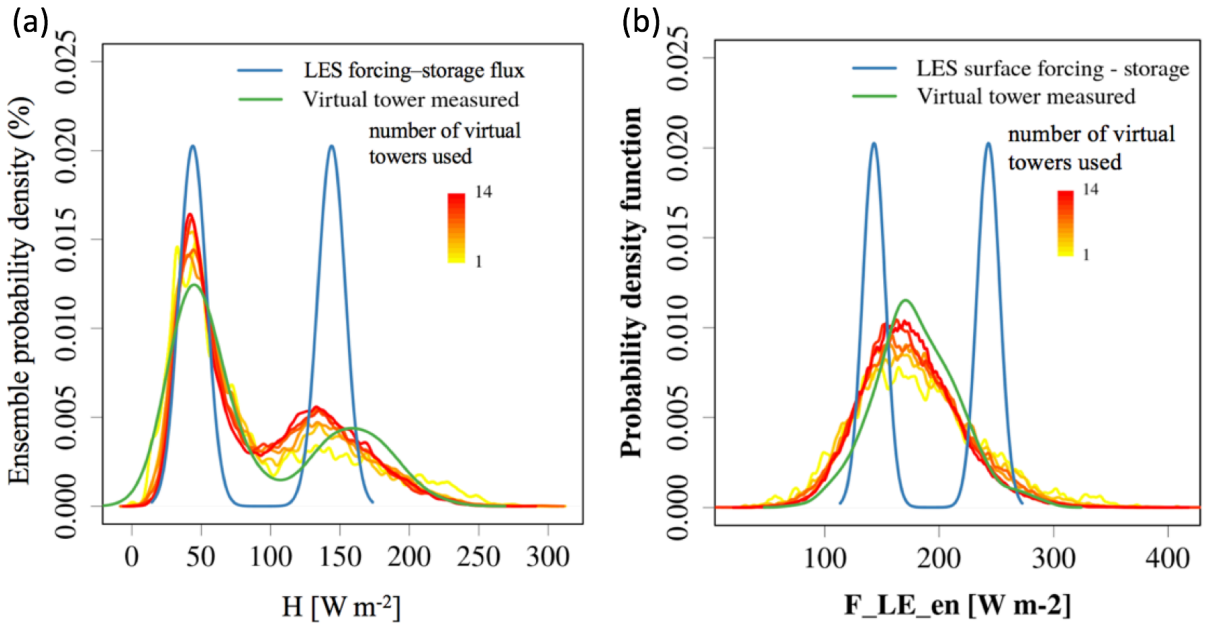


Figure 27: 2000-fold ensemble probability density functions of ERF-projected sensible heat and latent heat fluxes with 1 to 14 virtual towers. Adding more towers improved the reproduction of warm, strong convective patches. Even with few towers cold-wet patches can be detected more easily, and adding more towers does not substantially improve their reproduction.

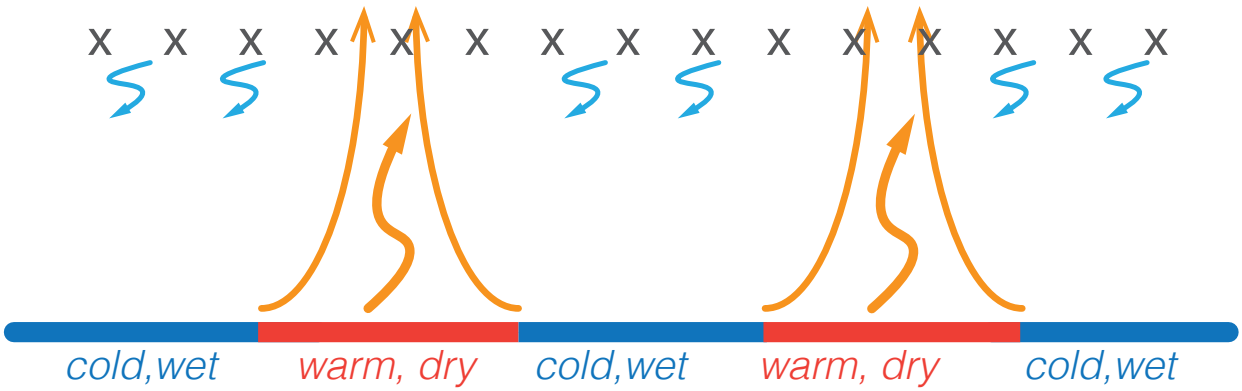
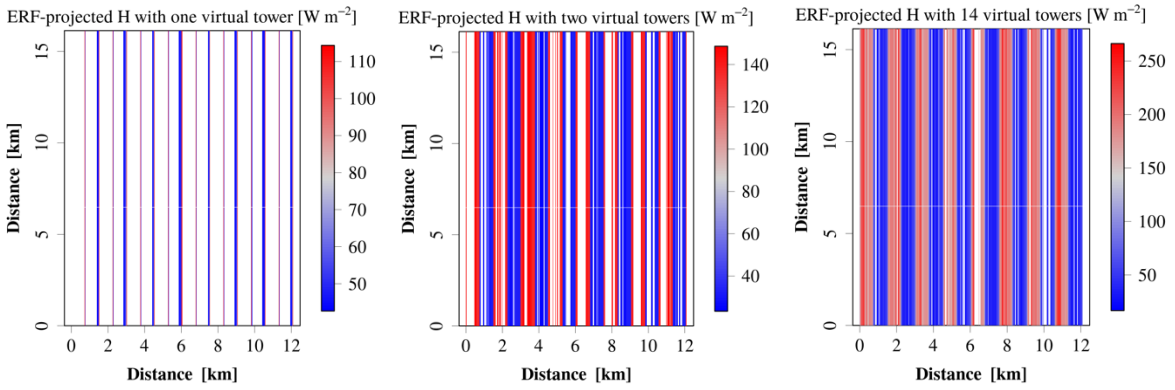


Figure 28: Schematic plot for atmospheric convection skewness. In the updraft branch (orange lines) of the mesoscale eddies, convective eddies gather more energy to overcome gravity, and become stronger and spatially-smaller as they go up. The weaker downward flow (blue lines) needs more area than the stronger upward flow so they compensate each other according to mass conservation. The corresponding spatial distribution of atmospheric fields is skewed towards weaker, cooler subsidence areas (as opposed to stronger, warmer convective areas), even though cold/warm surface forcing (real-world or virtual) are uniformly distributed in space (red and dark blue lines). Because the EC tower (black crosses) is operating in this atmospheric field, its measurement is inherently "biased" to capture more of the weaker downward flow as opposed to the stronger upward flow.



a **b** **c**

Figure 29 ERF-projected sensible heat fluxes with (a) one virtual tower, (b) two virtual towers, and (c) 14 towers. White areas are gaps that cannot be reproduced by ERF because their physical properties exceed the range of the observation dataset. The spatial coverage increases from 20% with one virtual tower, over 62% with two towers, to 89% with 14 towers.

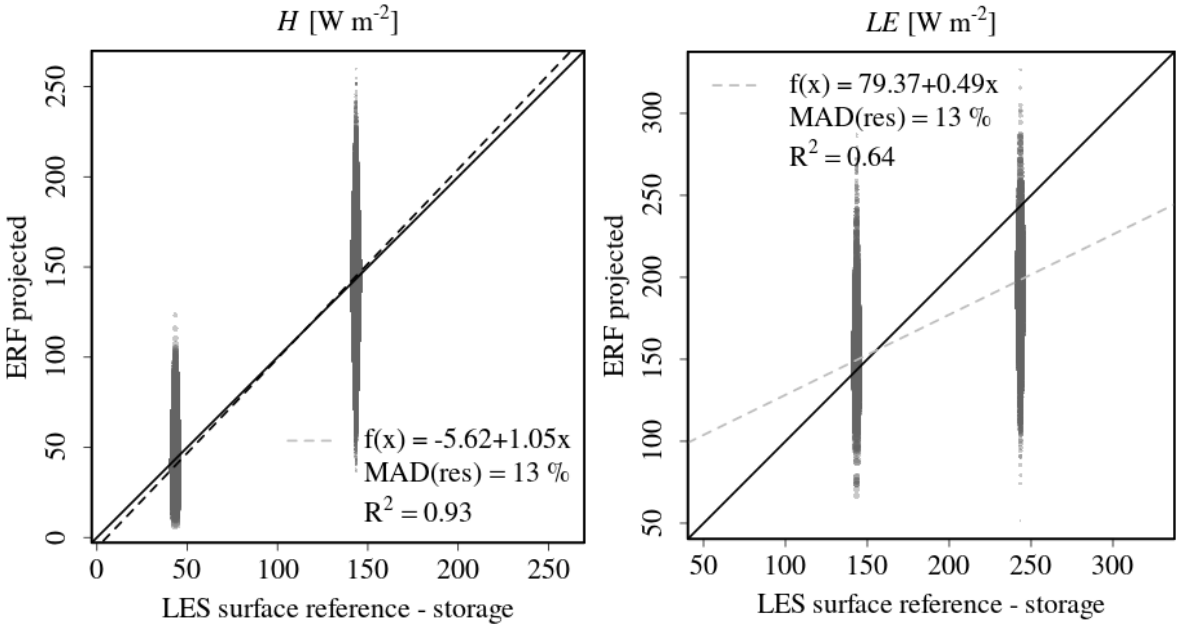


Figure 30 Scatterplot for ensemble ERF-projected sensible heat flux and latent heat produced from 14 virtual tower observations and LES surface reference subtracting storage flux. Solid black line is 1:1 line, and dashed black line is the fitted relationship.

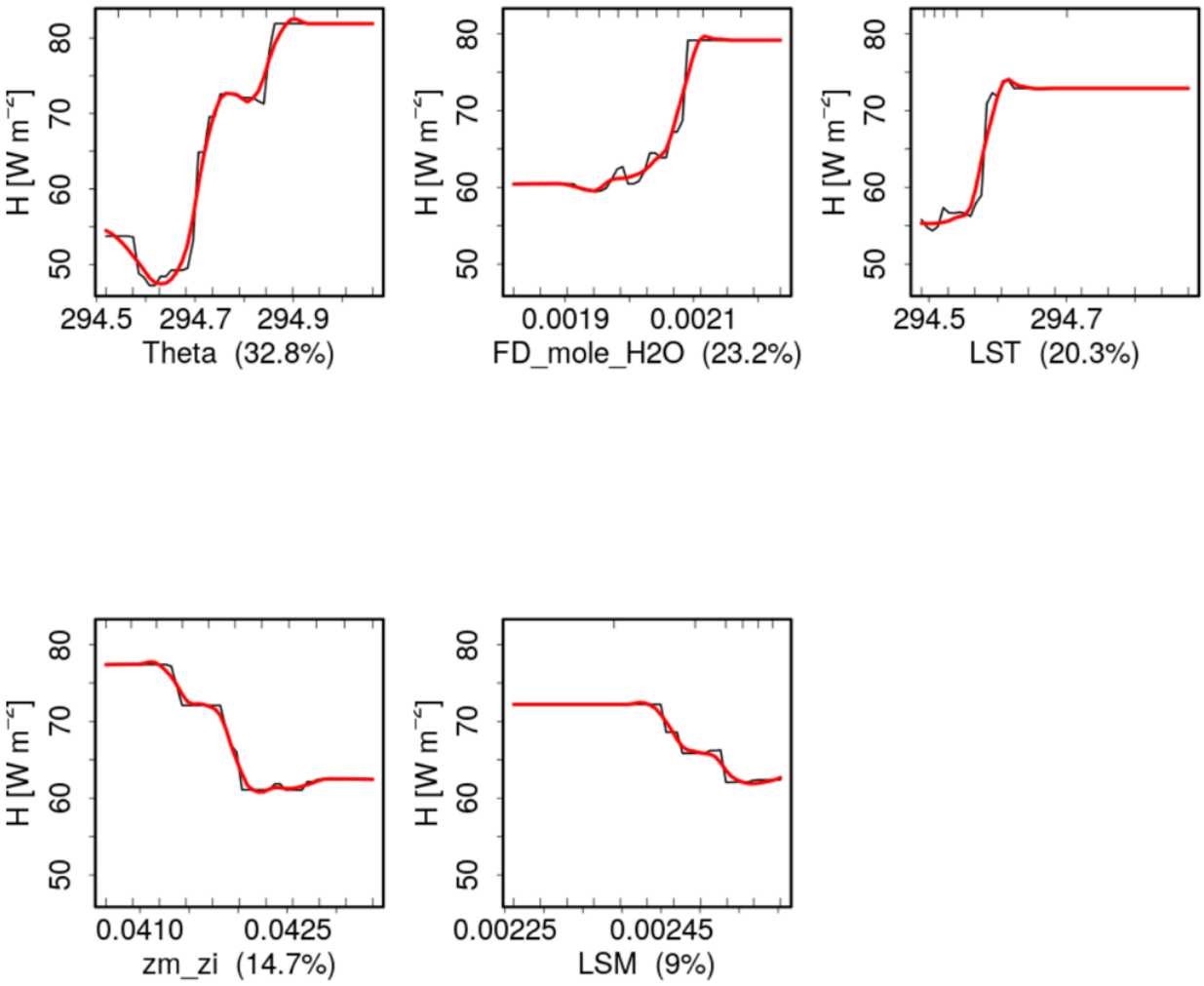


Figure 31 Extracted environmental response functions when 14 virtual towers are used. The fitted sensible heat flux (H) is shown on the ordinate. The abscissa show: potential temperature measured by the virtual tower (theta), water vapor dry mole fraction (FD_mole_H2O), land surface temperature (LST), relative measurement height in the boundary layer (zm_zi), land surface moisture (LSM). The extracted relationships are consistent with flux-gradient relationships in terms of relationship with LST, FD_mole_H2O, LSM. The positive relationship between theta and fitted H agrees our understanding that the air is usually warmed through large sensible heat flux and long wave radiation, especially over dry areas. The extracted relationship with zm_zi is consistent with the effect from vertical flux divergence.

5 Conclusions and Future Perspectives

5.1 Conclusions

My dissertation has extended the Environmental Response Function (ERF) approach to solve the two main challenges, energy imbalance and location bias, in observing net surface-atmosphere exchange from eddy-covariance flux towers. ERF achieves to decompose fluxes by utilizing the footprint variation over heterogeneous surface at high spatiotemporal resolution, relating the flux response to key environmental drivers, and projecting fluxes in space and time. Compared to traditional spatio-temporal eddy-covariance by adding more towers and previous upscaling methods, ERF has advantages in reducing observation requirements, resolving the missing sources in energy budget, and enabling flux decomposition at fine spatio-temporal resolution.

Chapter 2 demonstrated applicability of the ERF approach to map heat, water vapor, and CO_2 fluxes from an eddy-covariance tower to the regional scale of 20×20 km^2. In the real application, ERF enables to improve spatial coverage from tower footprint, <10% of the regional scale spatial domain, to >70%, >60%, and >50% over 5×5 km^2, 10×10 km^2, and 20×20 km^2, separately. The ERF-projected fluxes of sensible heat and latent heat and CO_2 fluxes 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%, respectively). 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 both independent observations and previous upscaling methods. 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.

Chapter 3 proved ERF-VCV is useful for retrieving the volume-controlled net surface atmosphere exchange (NSAE), when using AmeriFlux Park Falls WLEF tall tower in North Wisconsin, USA during July and August 2014. This retrieval is achieved by resolving the storage flux, vertical turbulent, and vertical advection fluxes, which are not easily measured. ERF approach is useful for mapping heat and CO_2 fluxes from an eddy-covariance tower footprint-variable representation to a fixed-coordinate representation at regional scale. This aids reducing the location bias typically incurred from single-location vertical turbulent flux as well as single-profile storage flux measurements, and the influence of vertical advection. Particularly, in this study, storage flux did not converge to zero as often assumed and had significant diurnal cycle, and should thus not be ignored, e.g. when comparing eddy covariance measurements with earth system models (ESMs) at hourly or finer temporal resolution. Low frequency flux contributions were detected in this study, and inclusion in the ERF-VCV landscape-scale exchange of sensible and latent heat led to a 20.6% increase over the tower observations, thus provides a promising research direction for improving energy balance closure.

Chapter 4 further proved that ERF enables detecting and resolving mesoscale eddies, which contribute to 4.3% out of 13.2% eddy-covariance energy imbalance over 12 km×16 km in the LES experiment, with one single tower, compared to 5 towers using traditional spatio-temporal eddy-covariance method. This advance is achieved by using wavelet-decomposed fluxes, which considers high-frequency (sub-hourly) fluxes with long transporting scales, e.g. one hour or three hours.

In this experiment, evidence does show mesoscale eddies in latent heat flux (LE). LE mesoscale eddies are not generated by LE itself. Rather, the updraft and downdraft branches generated by sensible heat fluxes trigger the mesoscale eddies in LE field.

In addition, ERF rectifies the location bias introduced by atmospheric skewness: strong convective areas are sparsely distributed, while cool subsidence areas cover the majority of the space. But in this LES analysis, for the surface, warm and cold patches are evenly distributed. Convection and heterogeneity-induced mesoscale eddies tend to increase ratio of cold-wet patches to warm-dry patches observed by towers, driving towers to under-sample warm-dry area. ERF rectifies atmospheric-skewed location bias by retrieving more warm areas in mapping surface atmosphere exchanges, while traditional spatio-temporal eddy-covariance using more towers cannot either detect or correct this atmospheric-skewed location bias.

The theoretical soundness and statistical significance of ERF approach are proved by both the very good agreement between ERF-projection and known LES surface reference, and the physically reasonable extracted response functions. ERF enables rectifying location bias by improving flux coverage over the regional domain from <1% tower footprint area, to 20%, 62%, and 89%, respectively, with one, two and 14 virtual towers.

Potential improvements will address the state-space gaps in the projected grids, assumption of spatially homogeneous meteorological drivers and coarse spatial-temporal resolution in surface property drivers. With spatially explicit net radiation, ground heat flux observations, and flux measurements from multiple flux towers and airborne eddy-covariance, the validation of ERF and the potential for improving the energy balance closure problem can be further studied from intensive observations.

ERF will be the next-generation eddy-covariance flux calculation algorithm that enhances the use and value of data from multiple flux networks. Perspective advances in the field will include

developing operative ERF routine among flux tower networks across north America (Sect. 5.2), providing next-generation benchmark for Earth system modelling (Sect. 5.3), and benchmarking for remote sensing data in data comparison, validation, and complementation (Sect. 5.4), advancing the study of boundary layer development over heterogeneous surface (Sect. 5.5).

5.2 Future expansion of ERF-VCV

A number of projects are underway to expand the ERF-VCV approach to make eddy flux observations a network-wide operative tool. NEON Surface-Atmosphere Exchange team led by Dr. Stefan Metzger, Dr. Trevor Keenan at the DOE LBNL, and I are targeting to expand the application of ERF-VCV from one single tower to flux network-wide across North America, through the scope of NEON Operations, NCAR Advanced Study Program fellowship application, and proposals to be submitted to the ROSES-ECOSTRESS call. Fundamentally, the goal is to move eddy-covariance flux observations from a micrometeorological tool, to one that is directly useful for regional to continental scale ecosystem ecology. The deliverable is to provide open-source, network-scalable ERF-VCV software as part of the community-developed eddy-covariance software 'eddy4R' in user-friendly R-packages in platform-independent Docker images (Metzger et al., 2017). This tool will automatically generate flux maps and response surfaces, e.g. around the 47 NEON eddy-covariance sites, at sub-100 m spatial and sub-hourly temporal resolution. Both data and routines will be available and sharable to all eddy-covariance users in the community. We believe developing and deploying an operational and automatic ERF-VCV routine among the entire networks can provide an end-to-end solution for location bias and energy imbalance in eddy-covariance observations, and thereby benefit all eddy-covariance data products users in the research community.

5.3 Prospect for model-data fusion

Uncertainty in terrestrial carbon fluxes simulated in ESMs can lead to uncertainty in future atmospheric CO_2 concentrations as high as several hundred ppmv and global mean temperature as large as several centigrade, given the same anthropogenic forcing. There are a number of sources of uncertainty that drive the divergent carbon fluxes in ESMs, but significant improvements are possible when confronting models with contemporary ecosystem observations, such as the eddy covariance data.

However, to date virtually all model-data inter-comparisons using eddy-covariance data have neglected uncertainties from footprint bias and energy balance non-closure. ESMs represent regular gridded cells (10^2 km^2–10^4 km^2), while eddy-covariance observations represent a much smaller footprint on the order of 10^{-1}–10^1 km^2 around the tower, and this footprint is temporally varying. The problems are well known, yet unified solutions remain elusive. Flux footprint rectification has been applied in limited domains (Desai et al., 2008; Wang et al., 2006; Xiao et al., 2014). Any transient bias that occurs from changes in sampled characteristics with time will bias ecological inference and model-data comparison, thus corrupting model improvement. Essentially, observations and modelling currently coexist on two completely different spatial scales, requiring some treatment of the observations prior to informing models.

The application of ERF in model-data fusion can generate both novel data products and novel understanding. As far as products are concerned, ERF can produce the hourly net ecosystem exchange of energy, water vapor, and especially CO_2 and even CH_4 flux maps over the regional scale (20×20 km^2) around tower sites. This can provide an unprecedented advantage compared to existing strategies for model-data fusion: Not only an expected value, but the entire probability density function over a model cell is quantified at hourly resolution. Utilizing the full probability density functions of carbon and water vapor fluxes from ERF products can certainly improve model parameter estimation and the determination of parametric uncertainty.

Moreover, the spatio-temporally decomposed CO_2 and CH_4 fluxes for particular land cover types enables directly compare fluxes from these land cover/vegetation types and evaluate model performance for the different plant functional types. As far as understanding goes, a set of extracted environmental response functions that provide the multi-dimensional non-linear sensitivity of the target flux to underlying drivers within the domain, providing a second mechanistic set of sensitivities for model comparison. 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 data bases, the public operational algorithm routines, and the developed software will rectify scale mismatch problem, thereby offering a network-wide observatory benchmark that can exert a revolutionary impact in model-data comparison, validation, and assimilation.

5.4 Prospect for remote sensing data in data comparison, validation and complementation

Fusing remote sensing observations with eddy-covariance tower measurements greatly amplifies the utility of both, providing novel understanding of the terrestrial response to climate change from local to global scales. However, the spatial scale mismatch exists between the two observational platforms: remotely sensed retrievals represent regular gridded cells (60 m–1 km), while eddy-covariance observations represent a temporally-varying footprint on the order of 10^{-1}–10^{1} km^2 around the tower. Hence, any observed dynamics or potential trends may be quite different only because of sampling a different source area (e.g., Griebel et al., 2016; Montaldo and Oren, 2016; Morin et al., 2017; Metzger 2017).

ERF can serve as a promising tool to address the aforementioned issue and hourly/half hourly scale-corrected next generation eddy-covariance data maps around North America flux tower networks, such as AmeriFlux and NEON. This will operatively accelerate flux networks' data products generation, standardization, and harmonization for advancing the understanding of terrestrial biosphere. Further, energy non-balance has also been found when complimenting remote sensing radiative observations with in-situ surface observations (L'Ecuyer et al. 2015; Wild et al., 2011; 2013).

ERF can serve as a promising tool to address the aforementioned issues and provide hourly/half hourly scale-corrected and energy-corrected next generation eddy-covariance data maps around North America flux tower networks, such as AmeriFlux and NEON.

The produced network-wide regional maps of surface-atmosphere exchanges of sensible heat (H), evapotranspiration (LE), net ecosystem exchange of CO_2 (NEE) will serve as a benchmark in validating and calibrating recently- and soon-to-be-launched satellite missions, e.g. ECOSTRESS, SMAP, Orbiting Carbon Observatory-2/3 (OCO-2/3), Table XX).

Table 1. Main satellite data products in multiple NASA missions that can be calibrated and/or validated using the produced near-real-time (NRT) eddy-covariance scale-corrected maps.

eddy-covariance NRT scale-corrected data products	Satellite data products that can be validated and/or calibrated against eddy-covariance data	Mission	Launch time	Temporal resolution	Spatial resolution
H	H	ECOSTRESS	2018-Apr	4 days	60 m

ET	ET	ECOSTRESS	2018-Apr	4 days	60 m
NEE	L4_C NEE	SMAP	2015	7 days	9 km
GPP	SIF-based GPP	OCO-2/3	2014/2018	16 days	2 km

The data bases, the public operational algorithm routines, developed software will provide a valuable prompt benchmark of surface-atmosphere exchange that can exert a revolutionary impact in catalyzing and facilitating data sharing, calibration/validation, model-data assimilation, and thereby benefit all eddy-covariance data products users in the research community. As a byproduct, extracted relationships unveil novel ecosystem responses from the information encoded but hidden in high temporal resolution in-situ observations. The relationships can be directly used to inform next-generation remote sensing algorithm and model structure and parameterization.

5.5 Prospect for the study in boundary layer development and atmospheric inverse models

A number of theories exist on how hypothetical or land surface variations drive planetary boundary layer (PBL) growth (Desai *et al*., 2006; Reen *et al.*, 2014), turbulence structure (Platis *et al.*, 2017), and cloud development (Gantner *et al.*, 2017), but current LES studies in heterogeneous terrain lack accurate estimate of surface conditions including surface fluxes. The ERF mapping in the prior section can provide realistic surface forcing for the LES model at high spatial resolution (100 m). The LES can then directly show how spatial variation in energy fluxes and land cover impart on the atmosphere.

Initial LES results (Fig. 32) of the domain during a case study analysis where ERF technique was used to map heterogeneous energy fluxes showed that heterogeneous terrain presents

slightly stronger and larger patterns (right panel) in turbulent structures (De Roo et al., 2014). We know from idealized numerical experiments using LES that this outcome depends on the degree of heterogeneity how the intensity of turbulent organized structures in modified.

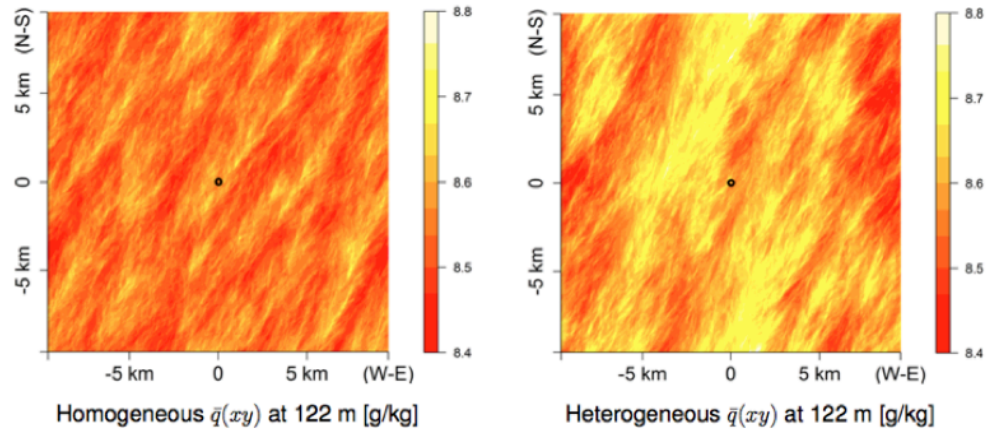


Figure 32 Comparison of LES water vapor field at 122 m in response to homogenous surface energy forcing (left) and hetereogenous ERF scaled energy forcing for the study domain. Tall tower is denoted by the dot. Courtesy of F. deRoo.

5.6 Future directions for ERF

The ERF-reproduced scale- and energy-corrected hourly NEE/LE/H flux grids, as well as the prospects discussed in Sects. 5.2–5.6, would allow ERF really push the boundary of making eddy-covariance observations more useful for benchmarking models, remote sensing data, and advancing the study in land-atmosphere interactions, terrestrial response to climate change, and atmospheric boundary layer development, thereby benefit all the eddy-covariance data users in the community.

5.7 Reference

Desai, A. R., Davis, K. J., Senff, C. J., Ismail, S., Browell, E. V., Stauffer, D. R., & Reen, B. P. (2006). A case study on the effects of heterogeneous soil moisture on mesoscale boundary-layer structure in the Southern Great Plains, USA Part I: Simple prognostic

model. Boundary-Layer Meteorology, 119(2), 195-238. doi:10.1007/s10546-005-9024-6

Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., . . . Stauch, V. J. (2008). Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. Agricultural and Forest Meteorology, 148(6-7), 821-838. doi:10.1016/j.agrformet.2007.11.012

Griebel, A., Bennett, L. T., Metzen, D., Cleverly, J., Burba, G., & Arndt, S. K. (2016). Effects of inhomogeneities within the flux footprint on the interpretation of seasonal, annual, and interannual ecosystem carbon exchange. Agricultural and Forest Meteorology, 221, 50-60. doi:10.1016/j.agrformet.2016.02.002

L'Ecuyer, T. S., Beaudoing, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., . . . Hilburn, K. (2015). The Observed State of the Energy Budget in the Early Twenty-First Century. Journal of Climate, 28(21), 8319-8346. doi:10.1175/jcli-d-14-00556.1

Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., . . . Desai, A. R. (2017). eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5. Geoscientific Model Development, 10(9), 3189-3206. doi:10.5194/gmd-10-3189-2017

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. Journal of Geophysical Research: Biogeosciences, 111(G2). doi:10.1029/2005JG000093

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E. G., & König-Langlo, G. (2013). The global energy balance from a surface perspective. Climate Dynamics, 40(11), 3107-3134. doi:10.1007/s00382-012-1569-8

Xiao, J., Ollinger, S. V., Frolking, S., Hurtt, G. C., Hollinger, D. Y., Davis, K. J., . . . Suyker, A. E. (2014). Data-driven diagnostics of terrestrial carbon dynamics over North America. Agricultural and Forest Meteorology, 197, 142-157. doi:10.1016/j.agrformet.2014.06.013