Observing System Simulation Experiments Double Scientific Return of Surface-Atmosphere Synthesis

Stefan Metzger[1,2], David Durden[1], Sreenath Paleri[2], Matthias Sühring[3], Brian J. Butterworth[2], Christopher Florian[1], Matthias Mauder[4], David M. Plummer[5], Luise Wanner[4], Ke Xu[6], Ankur R. Desai[2]

[1]National Ecological Observatory Network Program, Battelle, 1685 38th Street, Boulder, CO 80301, USA
[2]Dept. of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, 1225 West Dayton Street, Madison, WI 53706, USA
[3]Institute of Meteorology and Climatology, Leibniz University Hannover, Herrenhäuser Straße 2, 30419 Hannover, Germany
[4]Institute of Meteorology and Climate Research - Atmospheric Environmental Research, Karlsruhe Institute of Technology, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany
[5]Dept. of Atmospheric Science, University of Wyoming-Laramie, 1000 E. University Ave., Laramie, WY 82071, USA
[6]Dept. of Climate and Space Sciences and Engineering, University of Michigan-Ann Arbor, 2455 Hayward St, Ann Arbor, MI 48109, USA

Corresponding author: Stefan Metzger (smetzger@battelleecology.org)

Key Points

- Integrative observing system design can multiply the scientific return of surface-atmosphere field measurements.
- Catalyzing numerical simulations and first-principles machine learning open up observing system simulation experiments to novel applications.
- Use cases include natural climate solutions, emission inventory validation, urban air quality, and industry leak detection.

Abstract

The observing system design of multi-disciplinary field measurements involves a variety of considerations on logistics, safety, and science objectives. Typically, this is done based on investigator intuition and designs of prior field measurements. However, there is potential for considerable increase in efficiency, safety, and scientific success by integrating numerical simulations in the design process. Here, we present a novel approach to observing system simulation experiments that aids surface-atmosphere synthesis at the interface of meso- and microscale meteorology. We used this approach to optimize the Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors 2019 (CHEESEHEAD19).

During pre-field simulation experiments, we considered the placement of 20 eddy-covariance flux towers, operations for 72 hours of low-altitude flux aircraft measurements, and integration of various remote sensing data products. High-resolution Large Eddy Simulations generated a super-sample of virtual ground, airborne, and satellite observations to explore two specific design hypotheses. We then analyzed these virtual observations through Environmental Response Functions to yield an optimal aircraft flight strategy for augmenting a stratified random flux tower network in combination with satellite retrievals.

We demonstrate how this novel approach doubled CHEESEHEAD19's ability to explore energy balance closure and spatial patterning science objectives while substantially simplifying logistics. Owing to its extensibility, the approach lends itself to optimize observing system designs also for natural climate solutions, emission inventory validation, urban air quality, industry leak detection and multi-species applications, among other use cases.

Plain Language Summary

Computer models allow us to inform societal decisions in many areas of life. For example, they can predict severe weather or the fate of life-sustaining environmental resources such as water, food and air quality. In turn, these computer models are rooted in field measurements, which are the foundation of our understanding how the earth's surface and atmosphere interact. However, the design of these field measurements is often limited to investigator intuition and prior examples. Here, we demonstrate how combining existing computer models and artificial

intelligence prior to installing any instrumentation can substantially enhance the design process. While an existing computer model simulates a number of possible measurement designs, artificial intelligence ranks each design for its ability to address project needs. We used this approach to determine the aircraft measurement strategy for a large field measurement project that links multiple science disciplines. This doubled the usefulness of the aircraft measurements for informing the next generation of surface-atmosphere computer models alongside tower and satellite measurements. The approach is extensible to optimize measurement designs also for natural climate solutions, emission inventory validation, urban air quality, and industry leak detection, among other applications.

1 Introduction

High-quality field data are the backbone of surface-atmosphere research. However, there are inevitable tradeoffs in any field measurement among cost, logistics, safety, and our ability to address science objectives. Most of the time, these tradeoffs are evaluated in a heuristic or haphazard approach, or at least with limited consideration of all possible options. Nevertheless, redundancy, experience, and good fortune usually save most field measurement Observing System Designs (OSDs) from failure. Inspired by Observing System Simulation Experiments (OSSEs) in the earth system sciences (Atlas et al., 2015; Hoffman & Atlas, 2016; Masutani et al., 2010) we contemplated whether this process could be improved. In particular, we note modern advances in conducting virtual experiments within high-resolution numerical simulations (NSs) of atmospheric turbulence (e.g., Steinfeld et al., 2007). We envisioned that such NSs could yield OSSEs that increase 'scientific return on funding investments', more effectively address field measurement objectives, and minimize problems that arise from safety, logistics, and cost.

Here, we derive a novel approach to OSSEs that aids surface-atmosphere synthesis at the interface of meso- and microscale meteorology. We then apply it to preparing field campaign resources for the "Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors 2019" (CHEESEHEAD19; Butterworth et al., 2020). At the time of this study, the CHEESEHEAD19 field measurement campaign was to be conducted in northern Wisconsin, United States of America, from June to October of 2019, with the overarching science objective to examine how the atmospheric boundary layer (ABL)

responds to spatial heterogeneity in the surface-atmospheric exchanges of heat and water. Further science objectives were to test whether resulting mesoscale atmospheric processes relate to the lack of energy balance closure frequently observed by eddy-covariance (EC) towers. Lastly, CHEESEHEAD19 sought to apply advanced analytics over a multi-scale set of observations to yield scale-aware, energy-balanced data products that help improve model representation of sub-grid processes. To that end, we wanted to harness the complementarity among various in-situ and remote-sensing measurement systems.

However, the joint utility of these measurement systems for addressing the science objectives was not well characterized prior to the field measurement. Moreover, their joint utility is highly sensitive to the OSD including placement, and the resulting information overlap in space and time (Figure 1). Consequently, CHEESEHEAD19 scientific return hinged on our ability to merge information among the perspectives of ground-based, airborne and space-borne measurements, and numerical models. Plentiful data that are insufficiently connected to infer meaning risks data deluge rather than the next interdisciplinary breakthrough. While advances in post-field data assimilation aim to rectify limited and variable information overlap statistically (Williams et al., 2009), only the careful OSD of the field measurements themselves can treat their root cause. We thus sought an approach that empowers making informed OSD choices for surface-atmosphere field measurements in advance of the experiment.

Simulation experiments involve asking what would happen in an imaginary scenario and trying to understand whether the predicted outcome is compatible with existing theory. This form of inquiry is not an invention of modern science, but can be traced back at least to the empirical thought experiments of ancient Greek philosophers (Brown & Fehige, 2019; Palmerino, 2018). Famously, Albert Einstein employed thought experiments for his theoretical generalizations, including in his works on special and general relativity (Norton, 1991). With the rise of NSs came the opportunity to increase the complexity and detail of thought experiments, such as how to design meteorological field measurements (e.g., Cortina & Calaf, 2017; Eddy, 1974; Gehrke et al., 2019). More frequently, however, these NSs were reserved for applications where real-world tests would have been impractical or impossible (e.g., Wiens et al., 2003). These NSs centered on prescribing and propagating a-priori knowledge, i.e. creating “data from knowledge”. As a

result, the findings often remained subject to strong methodological assumptions that could not necessarily be met by real-world applications. More recently, the advent of data-intensive scientific discovery promises to offset some of these limitations by providing computational facilities that aid the inference of "knowledge from data", including machine learning (Hey et al., 2009; Reichstein et al., 2019). We believe that ours is the first work that explicitly complements these paradigms of scientific knowledge creation for deriving surface-atmosphere OSDs at the interface of meso- and microscale meteorology.

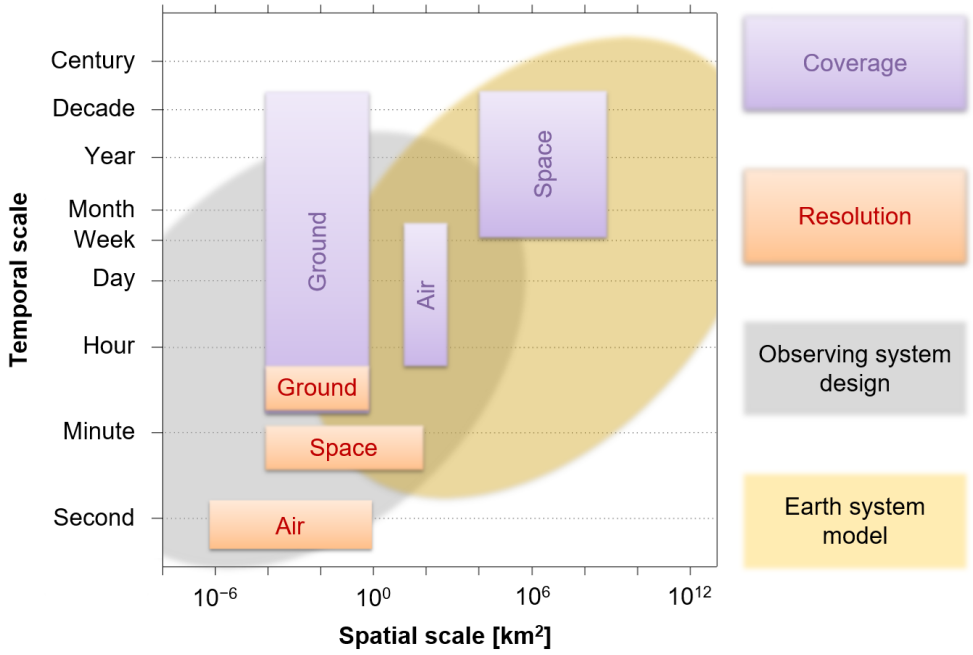


Figure 1. Space-time scope diagram for a surface-atmosphere synthesis observing system at the interface of meso- and microscale meteorology. The observing system consists of a hierarchy of ground-based (Ground), airborne (Air) and space-borne (Space) measurements, shown in relation to two principal approaches for scaling to an information continuum: pre-field observing system design and post-field data assimilation into earth system models. Modified after Metzger (2018).

Previous studies employed data-intensive scientific discovery for post-field OSD assessments (e.g., Chu et al., 2017; Koffi et al., 2013; Kumar et al., 2016; Loescher et al., 2014; Mahecha et al., 2017; Montanari et al., 2012; Villarreal et al., 2019). In comparison, one innovation of the presented approach is that it provides design information prior to testing OSDs in the field. To achieve this, we expanded on recent studies of atmospheric turbulence in Large Eddy Simulations (LES; Sühring et al., 2018; K. Xu et al., 2020). Specifically, we computationally simulated virtual observations over the CHEESEHEAD19 domain in decameter- and sub-second resolution. This "data from knowledge" feeds into a framework for data-intensive scientific discovery based on physics-guided Environmental Response Functions (ERFs; Metzger, 2018; Metzger, Junkermann, et al., 2013; K. Xu et al., 2017, 2018). The resulting explicitness promises unprecedented realism and process inference in comparison to existing pre-field OSSEs that leverage "knowledge from data" principles (Hargrove & Hoffman, 2004; Kaminski et al., 2012; Keller et al., 2008; Lauvaux et al., 2012; Lucas et al., 2015; J. Park & Kim, 2020; Sulkava et al., 2011; L. Zhang & Pu, 2010; Ziehn et al., 2016). In the following, we derive this LES-ERF OSSE approach (in short "LES-ERF" hereafter) using the design of CHEESEHEAD19 airborne flux measurements as a specific application example.

Airborne EC measurements have the particular benefit that they permit surface-atmosphere fluxes to be spatially resolved over a range of scales, from small, tower-sized flux footprints up to landscape scale. Thus, they build an important bridge among the differing scales of ground-based and space-borne measurements (Figure 1). Moreover, these kinds of measurements have the particular advantage that they can capture dispersive fluxes resulting from mesoscale atmospheric processes, which we hypothesize to be a main reason for the long-standing energy balance closure problem of tower-based eddy-covariance measurements (Margairaz et al., 2020; Matthias Mauder et al., 2020). In comparison to other ground-based and space-borne measurements, aircraft can be deployed quite flexibly in space and time. They thus provide a key to maximize the joint scientific return of harnessing complementarity among various in-situ and remote-sensing measurement systems. However, airborne field campaigns are very costly and cannot be conducted continuously. Therefore, thorough planning of the flight strategy is of great importance. Previous large-scale field campaigns with similar airborne flux measurement objectives include the First ISLSCP (International Satellite Land Surface Climatology Project)

Field Experiment (FIFE; P. J. Sellers et al., 1988), the Boreal Ecosystem-Atmosphere Study (BOREAS; P. Sellers et al., 1995), the Northern Hemisphere Climate Processes Land-Surface Experiment (NOPEX; Halldin et al., 1999), the Lindenberg Inhomogeneous Terrain - Fluxes between Atmosphere and Surface: a Long-term Study (LITFASS-98; F. Beyrich et al., 2002) and LITFASS-2003 (Frank Beyrich & Mengelkamp, 2006), MAtter fluxes in Grasslands of Inner Mongolia as influenced by stocking rate (MAGIM; Butterbach-Bahl et al., 2011), ScaleX (Wolf et al., 2017) and others. In these campaigns the flight strategies were mostly based on experience and expert knowledge. Considerations included random and systematic errors (D. H. Lenschow et al., 1994) and the source area (or "footprint"; Kaharabata et al., 1997; Schuepp et al., 1990), sometimes supported by analyzing land-cover maps using Geographic Information Systems (Metzger, Junkermann, et al., 2013). However, measurement errors and source areas not only depend on the flight track itself but also vary with atmospheric conditions, such as stability, wind speed, and wind direction. Moreover, focusing solely on measurement errors can be misleading in relation to heterogeneity-induced signals and result in erroneous conclusions (Sühring & Raasch, 2013).

The aim of this manuscript is to derive the theoretical background of LES-ERF, and to demonstrate its application to the CHEESEHEAD19 OSD with focus on the EC flight strategy. Specifically, in the following sections we test the study hypothesis that CHEESEHEAD19 scientific return is sensitive to LES-ERF optimization. Two accompanying design hypotheses relate this sensitivity to the choice of flight patterns and flight sequence. Sect. 2 introduces the methodology beginning with CHEESEHEAD19 (Sect. 2.1), and then derives LES-ERF, the design hypotheses and candidate OSDs (Sects. 2.2–2.5). Sect. 3 presents the LES-ERF results, beginning with LES (Sect. 3.1) and ERF (Sect. 3.2) specifics, then evaluating the design hypotheses for each candidate OSD (Sect. 3.3), and providing CHEESEHEAD19 field campaign resources (Sect. 3.4). Sect. 4 discusses these LES-ERF results, in light of the CHEESEHEAD19 OSD (Sect. 4.1), possible benefits for coordinated environmental observations in general (Sect. 4.2), and remaining challenges and future work (Sect. 4.3). Sect. 5 then summarizes our findings and provides an outlook.

2 Materials and methods

2.1 CHEESEHEAD19

The CHEESEHEAD19 study (Butterworth et al., 2020) was to be comprised of a four-month field measurement campaign to investigate how land surface heterogeneity influences energy balance closure. The energy balance closure problem refers to the situation, common in EC measurements, whereby downward energy from radiation and ground heat flux exceeds the measured upward energy from sensible and latent heat fluxes (Foken et al., 2011; Matthias Mauder et al., 2020). Previous studies have indicated that heterogeneity does impact energy balance (Stoy et al., 2013; Z. Xu et al., 2016). The CHEESEHEAD19 project proposed to evaluate the hypothesis that mesoscale features, driven by surface heterogeneity, are an important cause of energy balance non-closure (Charuchittipan et al., 2014; Foken et al., 2011; Gao et al., 2016; M. Mauder, Jegede, et al., 2007).

Due to a persistent mismatch between the scales of observations and models, the spatial and temporal scaling of surface fluxes is essential for evaluating theories on what happens within the sub-grid of atmospheric models, and how those feed back onto larger scale dynamics. Therefore, an additional science objective of the project was to use the unique, multi-scale set of observations to improve model representation of sub-grid processes and to assess the performance of ERFs for estimating the 'flux in a box' from the domain volume (Metzger, 2018; K. Xu et al., 2018).

The field measurement campaign was to be conducted within a 10 × 10 km domain of heterogeneous forest in northern Wisconsin, USA. It included patches of homogenous and mixed forests of evergreen, hardwood, and softwood deciduous trees, as well as grasses, wetlands, streams, and lakes with a characteristic surface length scale of 411±88 m (K. Xu et al., 2017). The domain was relatively flat, ranging from 455 m ASL in southwest to 500 m ASL in the northeast. Previous years' data from the study area showed that the summer months are characterized by light surface winds (typically < 5 m s^{-1}) coming predominately from the western hemisphere (180 – 360º) and daytime ABL heights ranging from 0.5 to 2.5 km above ground (mean of 1.5 km).

To measure fluxes (momentum, sensible heat, latent heat, CO_2) across the domain, 20 above-canopy EC towers were proposed to be deployed over a range of vegetation types (Figure 2). They

would measure all components of the energy balance, including net radiation, sensible and latent heat fluxes, and ground heat flux. The majority of towers were expected to be instrumented at 3 m – 32 m above ground, equaling 3 m – 15 m above the canopy depending on land cover. The exception would be the tall tower at the center of the domain which is an existing AmeriFlux supersite (US-PFa; A. Desai, 1996 -) that has been measuring fluxes at 30 m, 122 m, and 396 m above ground for the past 26 years (1995 - 2020; A. R. Desai et al., 2015).

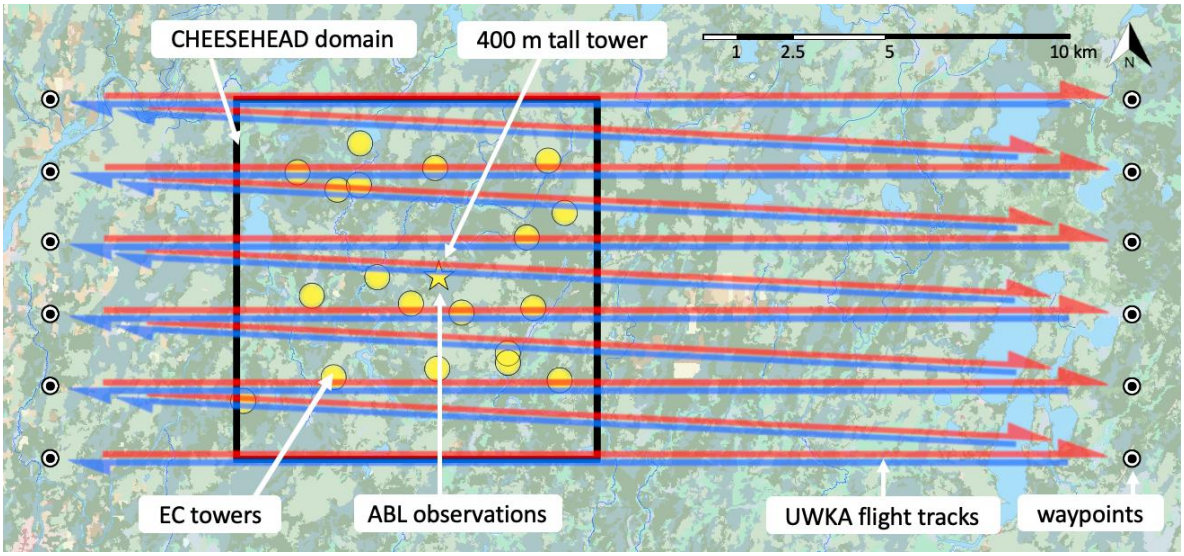


Figure 2. Set of candidate locations for EC towers, ABL measurements and UWKA candidate flight tracks (west-east pattern) with respect to the 10 × 10 km CHEESEHEAD19 study domain (black box; image: Google Earth). For a given flight track the UWKA would first fly outbound at 400 m above ground (red arrows) and return at 100 m above ground (blue arrows).

The project also proposed to deploy a suite of remote sensing instruments (lidar, radar, sodar, ceilometers, interferometers) for measuring the mesoscale atmospheric environment (profiles of wind, H_2O, temperature, aerosols, ABL height). Aircraft and space-borne remote sensing would be used to map surface characteristics that will be used for the data-driven scaling methods. This would include airborne hyperspectral imaging of the land surface properties. Additionally, a land surface temperature product was planned to be developed for the domain from a multi-sensor

fusion of in situ thermal drone and infrared camera imagery, ECOSTRESS, Landsat, VIIRS and/or GOES (e.g., P. Wu et al., 2013).

Aircraft measurements would be used to link the differing scales of ground-based and space-borne observations over the domain. Airborne EC fluxes (momentum, sensible heat, latent heat, CO_2) were to be measured with the University of Wyoming King Air (UWKA) during 24 research flights. The UWKA would also deploy an upward-pointing cloud lidar for estimating ABL height, and a downward pointing Raman lidar for providing a three-dimensional representation of air temperature and water vapor over the domain (D. Wu et al., 2016). During each research flight the UWKA would fly along 11 flight tracks (Figure 2). For a given flight track the UWKA would first fly outbound at 400 m above ground (Figure 2 red arrows) with emphasis on the lidar ABL observations. This arrangement would also allow the crew to visually ensure flight safety for the immediate return at 100 m above ground (Figure 2 blue arrows) with emphasis on the EC surface flux observations. Owing to the CHEESEHEAD19 science objectives we will focus on the 100 m EC surface flux flights in the following sections. A more complete description of the instruments deployed during CHEESEHEAD19 can be found in Butterworth et al. (2020).

2.2 LES-ERF observing system simulation experiment

Virtual atmospheres emulated in LES provide a controlled environment uniquely suited to disentangle surface-atmosphere feedbacks (e.g., Avissar & Schmidt, 1998; Kanda et al., 2004; Margairaz et al., 2020; Sühring et al., 2018; K. Xu et al., 2020). Our work on LES-ERF extends upon such previous use cases by simulating and analyzing candidate OSDs for real-world measurements in such virtual atmospheres.

LES-ERF employs value engineering principles (e.g., R. Park, 1998; Tohidi, 2011; Younker, 2003) to maximize the return on real-world measurement investments for addressing science objectives across traditional disciplinary boundaries. So long as we consider a single discipline, existing parameterizations often provide sufficient constraints to ensure meeting basic assumptions. For example, consider determining the height of an EC flux tower measurement as a function of roughness sublayer effects (e.g., Foken, 2017; Munger et al., 2012), atmospheric blending (e.g., Mahrt, 1996; Mason, 1988) and target source area (e.g., Chen et al., 2011;

Schmid, 1997). However, CHEESEHEAD19 relies on harnessing complementarity across disciplines, including ground-based, airborne, and space-borne measurements. These measurements operate on principally different space and time scales (Figure 1), so that scientific return hinges on our ability to join information not only across disciplines, but also across scales. For example, the spatial context of each measurement is a function of its horizontal and vertical placement, thus providing a mechanism to maximize information overlap. Yet, optimizing each measurement's utility for joint scientific inquiry is beyond the scope of discipline-specific parameterizations. Here, we propose the extensible LES-ERF approach that explicitly simulates the joint scientific return in response to different candidate OSDs for addressing user-defined design hypotheses.

In this study, we apply LES-ERF to derive an airborne EC flux flight strategy that augments a network of EC flux towers for optimally addressing CHEESEHEAD19 science objectives: relating surface-atmosphere feedbacks over a 10 × 10 km study domain to energy balance closure and space-time scaling (Sect. 2.1). A preconceived network of 20 continuously operating EC flux towers form CHEESEHEAD19's backbone for addressing these science objectives (Figure 2). Tower placement within the study domain followed a stratified random pattern, taking into account practical considerations including distance to road, suitable gaps in trees for a tower, USFS-owned land, etc. Individual towers were an average of 1.4 km from their nearest neighboring tower, and an average of 3.5 km from the tall tower. Here, we focus on a strategy for airborne EC flux measurements because they offer the greatest potential but also the greatest risk: (i) they are central to linking the different scales of ground-based and space-borne observations (Figure 1); (ii) their flexibility provides an accessible mechanism to maximize joint scientific return, and; (iii) their flight safety and cost warrant careful planning.

LES-ERF principally consists of three modules: (i) joining information across disciplines and measurements in ERFs, (ii) generating virtual measurements, e.g. in LES, and (iii) benchmarking candidate OSDs (Figure 3). ERFs augment expensive and thus sparse response observations (e.g., from tower and airborne EC) with inexpensive, abundant biophysical driver observations (e.g., from meteorological stations and satellites; Figure 3a). High-rate time-frequency decomposition and source area modeling facilitate mathematically rigorous data overlays among

these response and driver observations at minute- and decameter-scale. Machine learning then extracts a driver-response process model from the resulting space- and time-aligned dataset. Ultimately, this driver-response process model complements the properties of response and driver observations into a response data product. In the present example, the response data products are decameter-scale sensible heat flux maps, projected explicitly in space and time across the study domain. Sect. 2.5 provides specific implementation details for this module. To obtain virtual measurements ahead of the actual field measurement campaign, we used LES to create virtual atmospheres over the CHEESEHEAD19 domain for different synoptic conditions (Figure 3b). We then super-sampled these virtual atmospheres as observed by different yet simultaneous candidate OSDs. Sect. 2.4 details the specific implementation of this module. Each candidate OSD resulted in a separate set of virtual observations which we independently processed through the ERFs. Finally, we benchmarked each candidate OSD by comparing the flux maps that ERF reconstructed from the virtual observations alone to the original LES surface flux forcings (Figure 3c). To judge the ability to reproduce the LES reference we used three optimality criteria directly related to the CHEESEHEAD19 science objectives, each ranging 0–100%:

(CR1) Flux map spatial coverage, i.e. the percentage of grid cells across the study domain that ERF was able to reconstruct.

(CR2) Energy balance ratio;

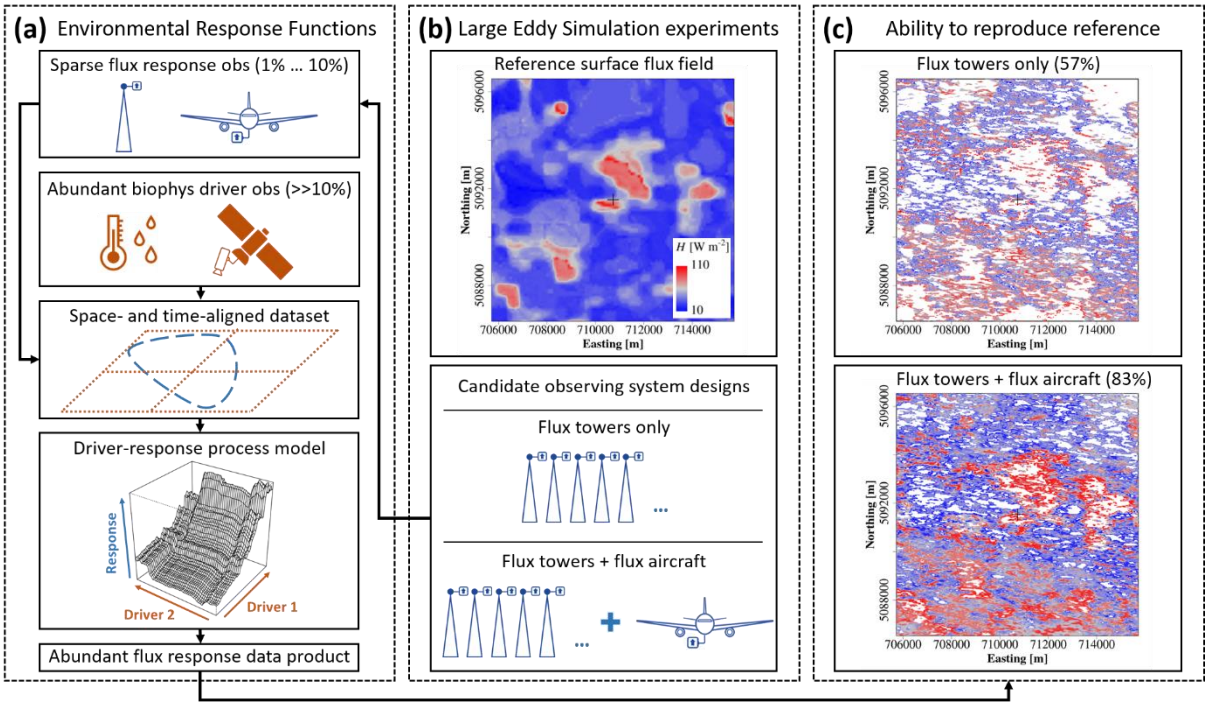
$$EBR = \frac{\langle \mathrm{F}_{H,ERF} \rangle + \langle \mathrm{F}_{LE,ERF} \rangle}{\langle \mathrm{F}_{H,LES} \rangle + \langle \mathrm{F}_{LE,LES} \rangle}; \quad (1)$$

with angle brackets indicating the horizontal average over all grid cells in the study domain, and F_H and F_{LE} the sensible and latent heat flux, respectively.

(CR3) Spatial patterning from point-wise Pearson correlation between the ERF reconstructed flux maps and the corresponding LES forcings.

We then used the arithmetic mean and standard deviation to aggregate CR1–CR3 across flight patterns, flight sequences, and ultimately among themselves into a single score (Sect. 3.3).

316



317

318 Figure 3. (a) Environmental Response Functions relate tower and airborne eddy-covariance flux
319 'responses' at very high space-time resolution to meteorological and surface drivers from
320 ground-based, airborne and space-borne measurements. We then use the extracted relationships
321 to reconstruct fluxes explicitly across the study domain. (b) To simulate different candidate
322 OSDs ahead of the actual field measurement campaign, we used LES to obtain observations in
323 virtual atmospheres over the CHEESEHEAD19 domain. (c): We then benchmarked the different
324 OSD candidates against their ability to reproduce the LES reference in the form of flux grids that
325 ERF reconstructed from the virtual observations alone. Sect. 2.2 provides additional detail.
326 Modified after Butterworth et al. (2020), copyright (2020), with permission from the American
327 Meteorological Society to reuse under the CC BY 4.0 license
328 (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cc-by-4.0).

329

2.3 Design hypotheses and candidate observing system designs

We used LES-ERF to determine the sensitivity of the optimality criteria in Sect. 2.2 in response to two specific design hypotheses. These design hypotheses define the trade space between CHEESEHEAD19 science objectives, flight time constraints, and straightforward flight operation. Related to flight pattern, we hypothesize that (H1) it is critical for airborne EC to measure perpendicular to the prevailing wind (e.g., Petty, 2020). And related to flight sequence, we hypothesize that (H2) within the flight time of a single research flight, it is more informative to fly a finely spaced pattern once compared to repeating a coarsely spaced pattern multiple times. To explore H1 and H2 we created candidate OSDs in an LES (Sect. 2.4), consisting of a virtual EC flux tower network in combination with virtual airborne EC flight patterns.

The virtual EC tower network formed the backbone of each candidate OSD, and its horizontal distribution corresponded to the CHEESEHEAD19 stratified random grid pattern. Nineteen virtual towers performed EC time-series measurements at 49 m above ground, i.e. 26±13 m higher compared to the real towers. The virtual AmeriFlux supersite tower at the center of the study domain measured at 49 m, 112 m, and 371 m above ground, i.e. −6±17 m lower compared to the real tower. These choices were a compromise among realism, bounding LES computational expense (10 m vertical resolution), and keeping the LES sub-grid fluxes acceptably small (<1%) as suggested by Schröter et al. (2000), which will not be captured by the virtual EC flux computation. We analyzed 2 h of data for each of the 22 virtual tower-level combinations, or 44 hours in total.

The virtual aircraft conducted EC space-series measurements along grid flight patterns at 100 m above ground, identical to the measurement height proposed for the real aircraft. The grid consisted of 11 flight tracks each 25 km long, including 6 parallel flight tracks 2 km horizontally spaced from each other, and 5 diagonal flight tracks in between (Figure 2 blue arrows). To create the dataset for assessing H1 we formed the virtual flight patterns by letting multiple aircraft fly simultaneous grids along four different azimuth angles of the parallel tracks: east-west (E-W), north-south (N-S), southwest-northeast (SW-NE), and south-southwest-north-northeast (SSW-NNE). Here, the term *flight pattern* refers to a set of georeferenced waypoints. To assess H2 we further considered three permutations of the *flight sequence*, i.e. the order in which the waypoints

of a given pattern are flown. (i) *Alternating* refers to flying a parallel track, then the downwind diagonal track, then the downwind parallel track, and so forth. (ii) *Outbound* refers to first completing all parallel tracks in downwind order, then flying back to the first parallel track and completing all diagonal tracks in downwind order. (iii) *Return* refers to first completing all parallel tracks in downwind order and then completing all diagonal tracks in upwind (return) order. To summarize, the virtual airborne EC dataset consisted of 3 flight sequences × 4 flight patterns × 11 flight tracks, or a total of 132 analyzed flight tracks spanning 3,300 km of virtual airborne EC data.

Based on this super-sample we evaluated 13 candidate OSDs. Applying LES-ERF to 44 site-hours of data from the virtual EC tower network alone provided the baseline OSD. Combining data from the virtual EC tower network with one of the 3 flight sequences × 4 flight patterns = 12 airborne EC combinations provided 12 alternative OSDs. Each of the alternative OSDs consisted of 44 site-hours virtual tower EC data and 11 flight tracks × 25 km = 275 km virtual airborne EC data. This configuration allows us to evaluate the change in the optimality criteria (Sect. 2.2) for each of the 12 joint tower and aircraft alternative OSDs relative to the tower-only baseline OSD.

2.4 Large Eddy Simulations

We used the Parallelized LES Model PALM (Maronga et al., 2020; Maronga et al., 2015), revision 4007 to simulate the atmosphere over the CHEESEHEAD19 domain. PALM solves the non-hydrostatic incompressible Boussinesq-approximated equations. We used the turbulent kinetic energy scheme of James W. Deardorff (1980) for the sub-grid model, a fifth-order scheme (Wicker & Skamarock, 2002) to discretize the advection terms, and a third-order Runge–Kutta scheme by Williamson (1980) for the time integration.

The aim of the simulation was to optimize OSDs for the upcoming field measurement campaign, meaning that the surface and atmospheric conditions were unknown. Hence, we simulated a day with a well-developed summertime continental ABL on 2011-08-12, which is a typical situation for that region during the scheduled field measurement period. We considered the model surface to be flat, and as surface forcing we prescribe time-dependent, heterogeneous sensible and latent heat flux grids that Metzger, Xu, et al. (2013) have previously determined for this day using

ERF. In an intermediary step we downscaled the original heat flux grids from 100 m to 25 m horizontal grid spacing and from 1 hour to LES time step, and filled data gaps with the horizontally averaged flux. This approach provides a straightforward surface flux benchmark for evaluating LES-ERF results, and forgoes the extensive data requirements of a coupled land surface model that would be difficult to fulfill prior to the actual field measurements. We then applied Monin–Obukhov similarity theory locally between the surface and the first vertical grid level as surface boundary condition for the momentum equations. During the pre-field stage, information on forest size and patch distribution was insufficient to use a plant-canopy model for reliably describing momentum drag. Hence, we set a horizontally homogeneous roughness length of 2.0 m in the simulations, because significant parts of the measurement site and its surroundings consist of forests. We then applied cyclic conditions at the lateral boundaries, and provided initial vertical profiles of the horizontal wind components, potential temperature and water vapor mixing ratio from nearby radiosonde observations (Green Bay Observations, Station ID 72645; ~100 km to the south east of the study domain). We assumed the observed westerly wind within the free-atmosphere to be in geostrophic equilibrium and steady state, and thus prescribed vertically constant profiles of the geostrophic wind components.

With this setup, we simulated a $30 \times 30 \times 3$ km^3 domain in x-, y-, and z-direction, respectively, centered over the 10×10 km CHEESEHEAD19 domain. The grid spacing was 25 m in the horizontal directions and 10 m in the vertical direction. The simulation ran for five hours (0700 – 1200 CST), of which the first three hours were model spin-up time (0700 – 1000 CST). During the final two hours (1000 – 1200 CST) we took virtual tower and aircraft measurements. At each virtual EC tower location, a virtual sensor at 49 m above ground sampled time series of potential temperature, mixing ratio, and vertical wind at the LES time step of 0.4 s. For each aircraft measurement, a similar virtual sensor moved along a predefined flight track at 100 m above ground with at a ground speed of 82 m s^{-1}.

2.5 Environmental Response Functions

ERF incorporates spectral averaging, source area modeling, and machine-learning-based space-time explicit ensembling of environmental observations (Metzger, Junkermann, et al., 2013). Here we used ERF to reproduce the LES surface flux forcing from virtual EC tower, EC aircraft,

and remote sensing observations (e.g., Serafimovich et al., 2018; K. Xu et al., 2017). These flux maps comply with observational assumptions that are not typically met from EC measurements alone, such as incorporation of mesoscale flows and spatial representativeness for the 10 × 10 km CHEESEHEAD19 target domain (Metzger, 2018; K. Xu et al., 2018; K. Xu et al., 2020).

ERF commenced with the high-rate time-frequency decomposed computation of the sensible and latent heat flux responses in the atmosphere. This step is based on the Morlet Wavelet, while assuming constant ambient pressure in the LES. Spectral averaging over the Wavelet cross-scalograms facilitated high temporal (tower: 1 minute) and spatial (aircraft: 100 m) resolution of the resulting fluxes (M. Mauder, Desjardins, et al., 2007). To ensure that tower and aircraft fluxes are comparable in their inclusion of long-wave mesoscale flows we applied a joint rectangular cutoff. The aircraft data limited the long-wave transport scales, with the 25 km flight tracks equating to a maximum transport scale of ~17 km expressible by the Wavelet cross-scalograms. We then applied Taylor's hypothesis (Taylor, 1915) with an average wind speed of 3–5 ms^{-1} to derive a corresponding tower cutoff scale of ~1 h for the tower. We time-matched the sensible and latent heat flux responses with the virtual observations of meteorological drivers consisting of potential temperature, water vapor dry mole fraction from mixing ratio, and relative measurement height in the ABL calculated from the potential temperature profile.

Next, we used source area modeling (Kljun et al., 2004; Metzger et al., 2012) to quantify the source area contributions to each 1 min tower and 100 m aircraft flux observation. The source area weights provided a linkage between the sensible and latent heat flux responses in the atmosphere and their spatially resolved drivers at the LES surface (available energy as a proxy for net radiation) and in the first vertical LES layer (land surface temperature and moisture as a proxy for remote-sensing observations). For simplicity, we used averages over the 2 h observation period for all spatially resolved drivers. The results are space- and time-aligned datasets consisting of the sensible and latent heat flux responses and their meteorological drivers in the LES atmosphere, and their spatially resolved drivers near the LES surface. The space-time aligned dataset for the baseline OSD (tower-only) thus consisted of 22 virtual tower-level combinations with 2 h of data each at 1 min output resolution = 2,640 observations. The space-time aligned dataset for each of the 12 alternative OSDs (tower + aircraft) additionally consisted

of 11 flight tracks with 25 km data each at 100 m output resolution = 2750 observations. This is the first application of ERF to combine flux response information across platforms, here flux tower and flux aircraft.

We then used boosted regression tree (BRT) machine learning to mine the information contained in the space-time aligned datasets. The results were individual ERF process models for each OSD, separately for the sensible and latent heat flux responses as a function of their meteorological and surface drivers. Overall, we built the driver-response model structure in accordance with first principles: an energy source, from which fluxes result in accordance to Fick's law of (turbulent) diffusion along temperature and water vapor gradients, modulated with distance from the exchange surface. For example, we expressed the sensible heat flux response as a space-time function of the vertical temperature gradient. BRT then solved for the turbulent diffusion coefficient as a space-time function of available energy, modulated by vertical flux divergence and the vertical humidity gradient.

In the final step we projected the space-time explicit heat flux response maps that the ERF process model yields when provided the full complement of space-time explicit drivers. Specifically, we provided the spatially distributed land surface temperature and moisture fields, the 2 h space-time median available energy across the 30 × 30 km domain, and the 2 h median meteorological drivers across all 20 virtual towers measuring at 49 m. This essentially equates to substituting the spatial information in the source areas with the distributed spatial information of land surface temperature and moisture fields akin to remote sensing. While it would have been possible to resolve the meteorological drivers temporally and hence also the resulting heat flux maps, we used the 2 h aggregates to streamline the overall analysis. We also limited the ERF projection to interpolate but not extrapolate outputs, i.e. to only populate grid cells with driver combinations in the range of the virtual measurements. By doing so, the resulting data coverage lets us directly estimate how well we sampled the domain for upscaling. In total, we trained and projected 78 ERF process models, consisting of two heat fluxes – sensible and latent heat – and 13 OSDs with 3 replicates each to constrain BRT tolerances.

3 Results

3.1 LES virtual experiments

As described in Sect. 2.4, the simulations were forced using ERF derived surface sensible and latent heat fluxes across the domain. Figure 4 shows the prescribed surface sensible and latent heat fluxes at different points in time which were used as lower boundary condition for the LES. The hourly input fluxes were interpolated in time to the current LES time step. Surface heterogeneities with distinct patches in the surface sensible and latent heat flux are visible at multiple scales that vary in time and among the latent and sensible fluxes as well. Distinct surface heterogeneity patches are maintained over the entire simulation period, representing particular landscape pattern within the field measurement site.

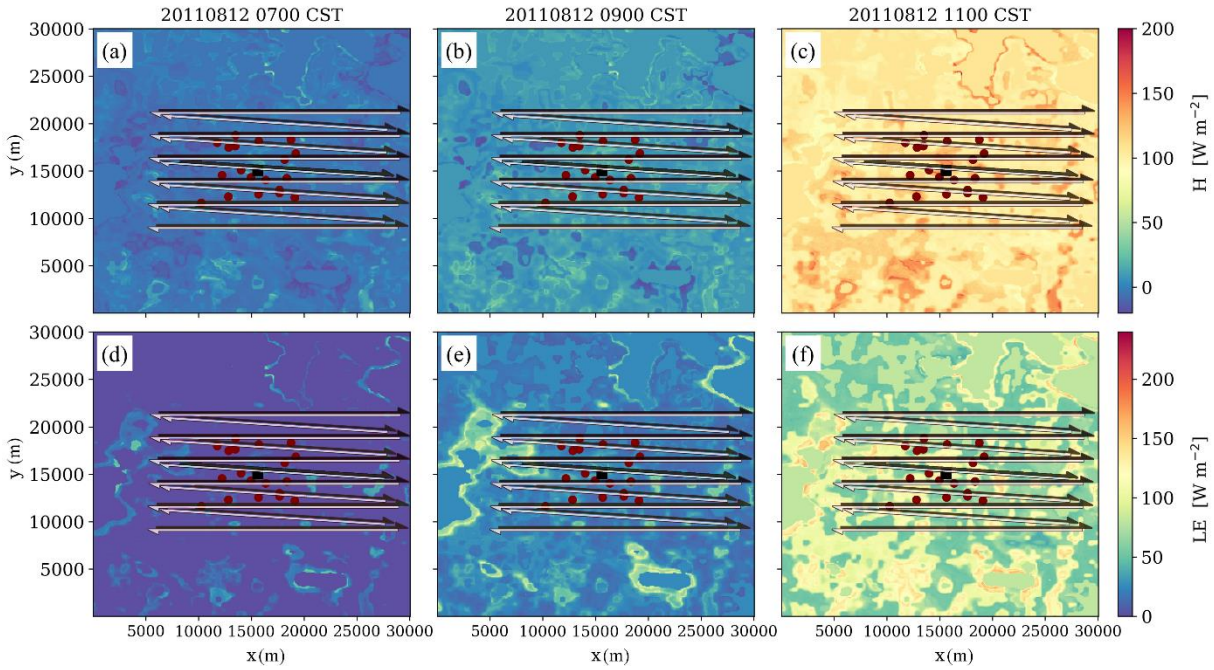


Figure 4. Time-sequence of (a) – (c) the spatially heterogeneous surface sensible heat flux and (d) – (f) latent heat flux prescribed as LES lower boundary conditions. Superimposed red dots indicate candidate EC tower locations, alongside UWKA candidate flight tracks (west-east pattern).

Figure 5 shows the domain-average initial and time-dependent vertical profiles of potential temperature, water vapor mixing ratio and wind speed. These explain the virtual setup and provide an overview of the ABL structure: The model was initialized with the early morning profiles of potential temperature and mixing ratio, and let to develop its own equilibrium for this design case. The profile of potential temperature indicates a vertically well mixed ABL which heats up during the course of the day. Due to the strong capping inversion the ABL grows only slowly and reaches a height of about 400 m around noon, which is relatively low for a summertime convective ABL. We discuss this in more detail in Sect. 4. The mixing ratio within the ABL also increases during the simulation due to the surface latent heat flux as well as due to entrainment of moist air from the above-lying free-atmosphere, which exhibits higher values of mixing ratio than in the ABL. The profiles of the wind components indicate northwesterly winds within the ABL during the morning hours, turning to westerlies later. Westerlies during the actual virtual measurement duration from late morning until noon are required to investigate the candidate OSDs from Sect. 2.3.

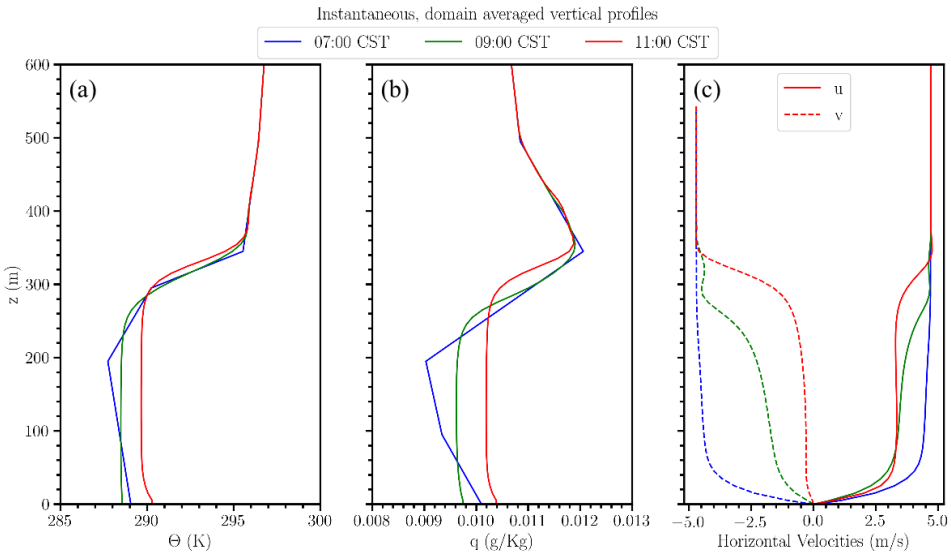


Figure 5. LES domain-averaged vertical profiles of (a) potential temperature, (b) water vapor mixing ratio and (c) horizontal wind velocities at different simulation times.

Figure 6 shows a horizontal cross-section of the instantaneous and time-averaged vertical wind component at a height of 100 m during the virtual measurement period at 11:00. Instantaneous up and downdrafts ranging from −2 m s^{-1} to 3 m s^{-1} can be observed. The up and downdrafts indicate elongated structures aligned with the mean-wind direction at this height level. Even though these organized structures are not stationary in space due to the slightly changing wind direction (see Figure 5), they still can be observed in the temporal average.

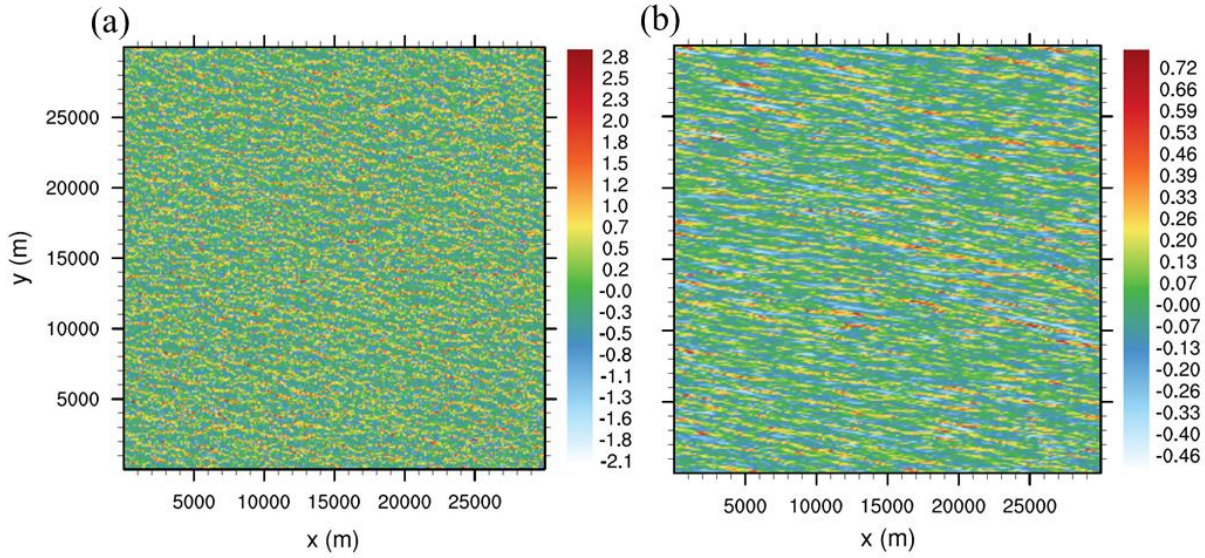


Figure 6. LES x-y cross section of (a) instantaneous and (b) 30-minute time-averaged vertical velocity at 100 m height at 11:00 CST simulation time.

Figure 7 shows vertical profiles of the domain-averaged sensible and latent heat flux. Both flux profiles display a similar shape, linearly decreasing with height and reaching a minimum at the ABL top. These negative heat fluxes indicate entrainment of warm and moist air from the inversion into the ABL. This is supported also by Figure 5, where the inversion layer exhibits a higher mixing ratio compared to the ABL. Figure 7 further shows that the subgrid-scale fluxes contribute less than 1–2 % to the total vertical transport at the measurement levels. This indicates that the vertical transport of heat and moisture is well resolved at these levels.

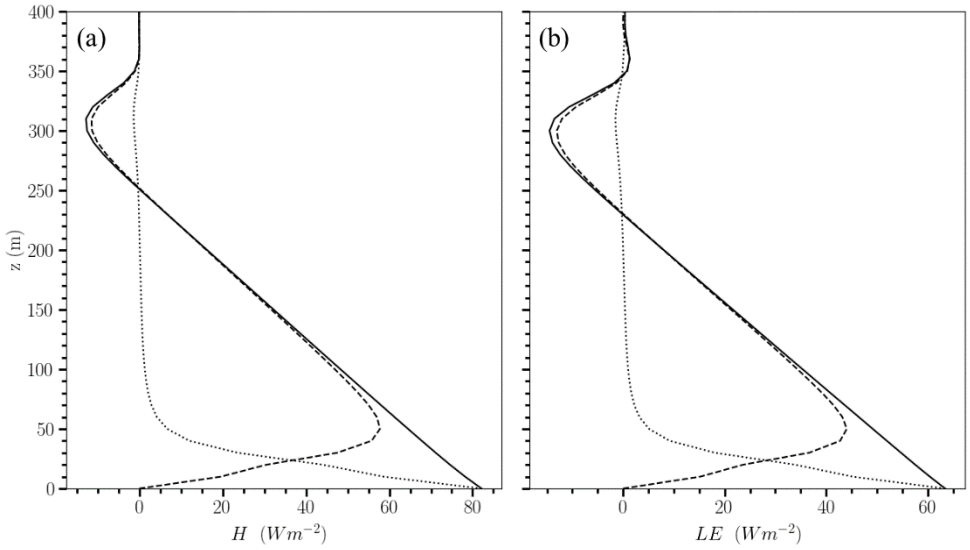
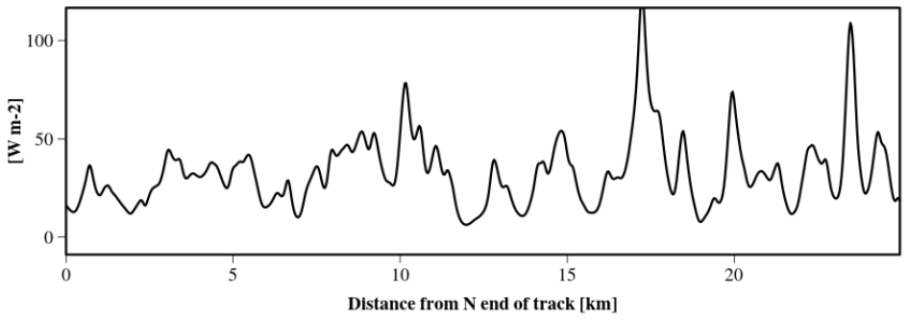


Figure 7. Domain averaged vertical flux profiles of (a) sensible heat and (b) latent heat at 11:00 CST simulation time. The solid lines show the total simulated fluxes, consisting of resolved fluxes (dashed lines) and sub-grid parameterized fluxes (dotted lines).

3.2 ERF retrievals

To create a space- and time-aligned data set (Figure 3a), ERF first calculates Wavelet-decomposed EC fluxes directly from the high-frequency raw data. This facilitates inclusion of longer transporting scales compared to traditional EC, as well as unprecedented spatial and temporal resolution of the resulting fluxes (Figure 8).

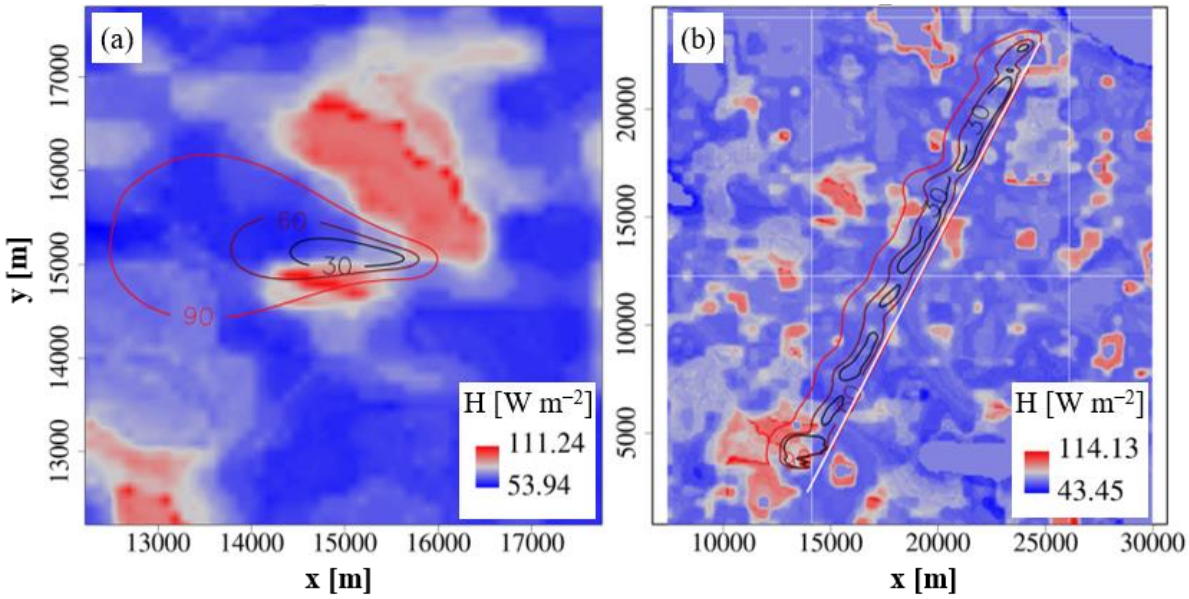


538 Figure 8. Space-resolved sensible heat flux from high-rate space-scale decomposition of virtual
539 airborne measurements.

540

541 Next, ERF relates the time- and space-resolved EC flux responses in the atmosphere to
542 biophysical drivers on the surface (Figure 9), such as land surface temperature and land surface
543 moisture. In the present application land surface temperature and land surface moisture are taken
544 from cross-sections at the vertical LES level closest to the surface. In real-world ERF
545 applications, these are substituted with space-borne and airborne remote sensing data products,
546 or reanalysis data (e.g., Serafimovich et al., 2018). This facilitates mathematically rigorous data
547 overlays among response and driver observations at minute- and decameter-scale. The result is a

548



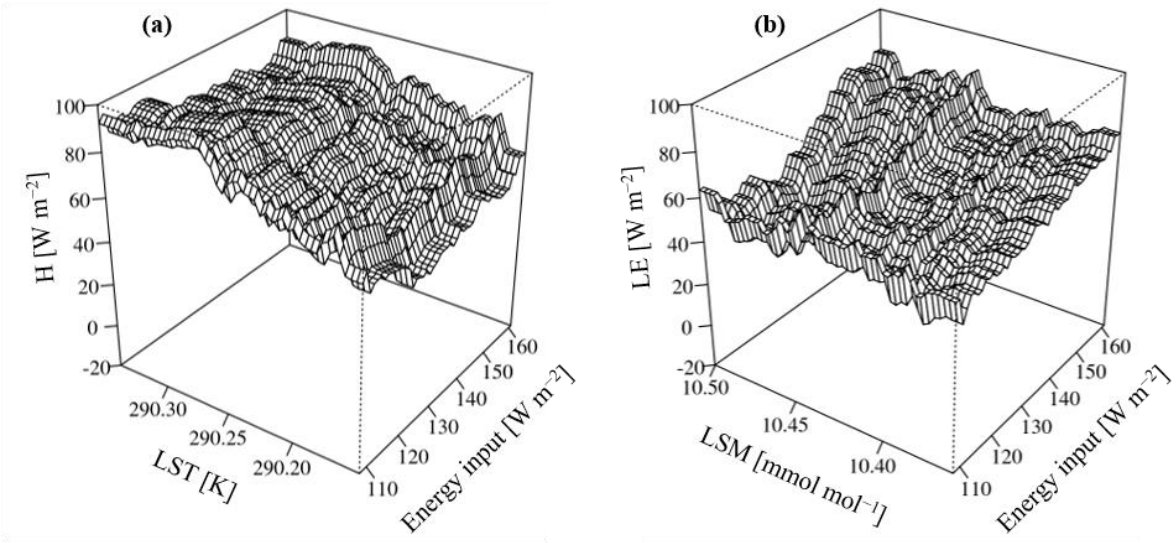
549

550 Figure 9. Source area modeling (30%, 60%, 90% contour lines) relates observations across
551 platforms and representations by determining the surface sources of the time- and space-resolved
552 EC fluxes. Here superimposed over the LES sensible heat flux surface forcing (a) Virtual
553 AmeriFlux supersite tower at the center of the study domain at 112 m measurement height. (b)
554 Virtual UWKA flight track at 100 m measurement height.

555 space-time aligned data set for each virtual EC tower and for each virtual EC flight track. Both,
556 the tower and airborne EC data sets comprise the same variables in identical units, and were
557 processed to ensure cross-platform compatibility and avoid biases (Sect. 2.5). This allows
558 combining the virtual tower EC results with corresponding virtual airborne EC results into a
559 single space-time aligned data set for each of the 12 alternative OSDs.

560 The ERF machine learning component then extracts a driver-response process model from the
561 resulting space- and time-aligned dataset. Here, we trained a total of 78 machine learning
562 models. These consisted of 13 candidate OSDs × 2 fluxes (sensible and latent heat) × 3 replicates
563 (to quantify stochastic uncertainty in the response data products). Figure 10 shows example
564 driver-response surfaces for sensible and latent heat flux as a function of their principal drivers,
565 energy input, land surface temperature and land surface moisture, respectively. This exemplifies
566 how the turbulent diffusion coefficient connects the drivers to the flux response within the
567 physics-guided ERF model structure. In Figure 10a the sensible heat flux increases primarily

568



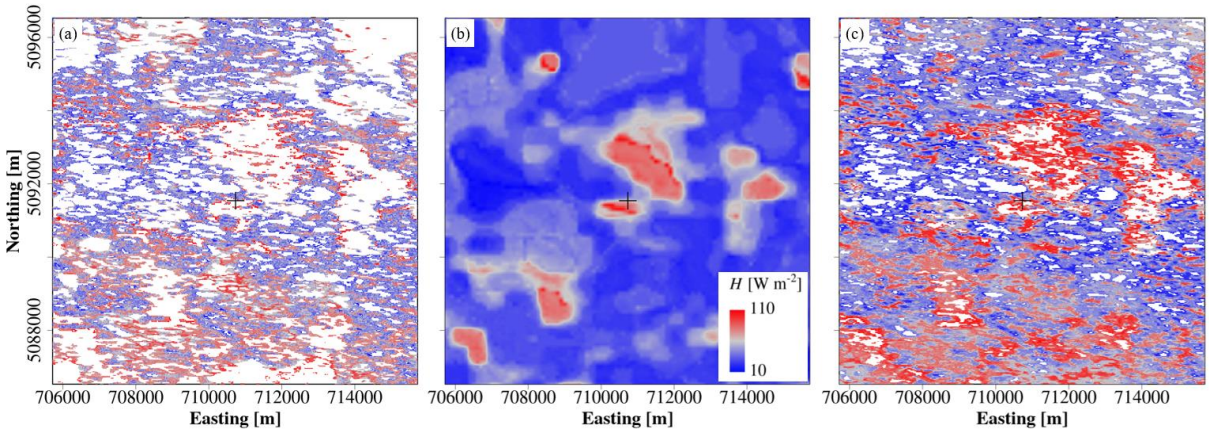
569

570 Figure 10. Example ERF response surfaces. (a) Sensible heat flux as a function of energy input
571 and land surface temperature (LST). (b) Latent heat flux as a function of energy input and land
572 surface moisture (LSM). For this visualization, all other drivers are kept at their median value.

573 with land surface temperature and secondarily with energy input. The relationship reaches a
574 plateau around 290.3 K which deviates from a one-dimensional, monotonic and linear gradient-
575 flux relationship, indicative of additional feedback processes. Conversely, in Figure 10b the
576 latent heat flux increases primarily with energy input and secondarily with land surface moisture,
577 with monotonic and approximately linear relationships across the range of drivers.

578 Ultimately, the physics-guided ERF driver-response process model complements the properties
579 of response and driver observations into a response data product. In the present example the
580 response data products are decameter-resolution sensible heat flux maps, projected explicitly in
581 space and time across the study domain to the median relative measurement height of the 49 m
582 towers (0.16 of the ABL height; Figure 11). We projected the flux maps for the tower-only
583 space-time aligned data set (baseline OSD; Figure 11a), and then separately for each of the 12
584 joint tower and aircraft space-time aligned data sets (alternative OSDs; Figure 11c). Now the
585 flux maps that ERF reconstructed from the virtual measurements alone can be compared to the
586 original LES surface flux forcings (Figure 11b).

587



588

589 Figure 11. Example ERF response data products: sensible heat flux maps independently derived
590 for (a) the tower-only space-time aligned data set and (c) for the joint tower and aircraft space-
591 time aligned data set, alongside (b) the LES reference surface flux field.

592

3.3 Evaluation of design hypotheses

The ERF-derived flux maps alongside the LES surface forcing in Figure 11 allow us to assess the design hypotheses (Sect. 2.3) as a function of the different candidate OSDs. For this purpose, we evaluated the change in the optimality criteria (Sect. 2.2) for each of the 12 joint tower and aircraft alternative OSDs relative to the tower-only baseline OSD. In the following Table 1 – Table 3 we performed all aggregations using arithmetic mean and standard deviation operators. In all cases the aggregations include two fluxes (sensible and latent heat) with three machine learning replicates each, plus additional aggregation as specified.

In response to the first design hypothesis H1 we address the question how critical it is for airborne EC to measure perpendicular to the prevailing wind. Table 1 shows the results for each optimality criterion as a function of the aircraft track angle on the mean wind direction, aggregated over all three flight sequences. We can see that track angles in the range of 90°±45° on the mean wind direction yield limited improvement in spatial coverage (23.3±1.8% – 25.6±0.1%) compared to wind-parallel patterns (0°; 20.9±1.9%). However, within the same range of track angles the improvement in energy balance ratio and spatial patterning approximately double to octuple (Table 1 green cells).

Table 1. Percent improvement of the joint tower and aircraft EC alternative OSDs relative to the tower-EC-only baseline OSD, aggregated over all three flight sequences. The results are shown as a function of optimality criterion (rows) and aircraft flying the grid pattern in various track angles on the mean wind direction (columns). Green cells highlight marked improvements that are further discussed in the text.

Optimality criterion	**All angles**	**0°**	**45°**	**60°**	**90°**
Spatial coverage	23.6±2.2	20.9±1.9	24.7±0.8	23.3±1.8	25.6±0.1
Energy balance ratio	6.8±5.3	1.7±1.4	6.4±5.3	6.4±4.7	12.8±3.1
Spatial patterning	23.2±11.7	13.7±9.2	34.6±3.3	26.2±6.8	18.3±15.2

The improvement in spatial patterning when adding wind-parallel flights to the tower network is limited to 13.7±9.2%, compared to 18.3±15.2 – 34.6±3.3% for adding flights with 45° – 90° aircraft track angle on the mean wind. Irrespective of the track angle, the observations along a flight track are never entirely independent from each other due to along- and cross-wind dispersion. For wind-parallel flights, Figure 12a indicates a high degree of source area overlap and thus self-correlation among the observations, resulting from strong along-wind dispersion along the flight track. In contrast, Figure 12b shows less overlapping source areas along the flight track of wind-perpendicular flights, with the comparatively weaker cross-wind dispersion now controlling the overlap. The latter strategy results in observations that capture more independent samples and spatial variability. Thus, the dominating mode of atmospheric dispersion with respect to the aircraft track angle helps to explain the differences in the spatial patterning optimality criterion. For our study setup we reach a critical overlap resulting from the combined effects of along- and cross-wind dispersion at track angles shallower than 90°±45° on the mean wind direction.

Furthermore, at the virtual aircraft flight height of 100 m the time-averaged vertical wind cross-section in Figure 6b shows organized structures that are elongated in the mean-wind direction. These organized structures are among the most promising leads to explain the frequently observed non-closure of the energy balance, in particular from tower EC measurements (Matthias Mauder et al., 2020). These structures consist of more spatially-expansive though weaker subsidence zones and more spatially-limited though stronger convection zones (Etling & Brown, 1993; Kanda et al., 2004; Donald H. Lenschow & Stankov, 1986; Moeng & Rotunno, 1990; Petty, 2020). So, when applied to aircraft EC measurements, wind-parallel flights are more likely to occur along the subsidence zones than along the convection zones. This helps explain why adding wind-parallel flights to the tower network yields only a limited improvement of the energy balance ratio criterion (1.7±1.4%). Conversely, wind-perpendicular flights trend to observe combinations of subsidence- and convection zones that approximately balances the atmospheric conservation of mass and energy. This explains the eightfold improvement (12.8±3.1%) when adding wind-perpendicular flights to the tower network compared to wind-parallel flights. In between these two extreme cases, adding the flights with 45° and 60° track

angles to the tower network still yields an approximately fourfold improvement (6.4±4.7% – 6.4±5.3%) over the wind-parallel flights.

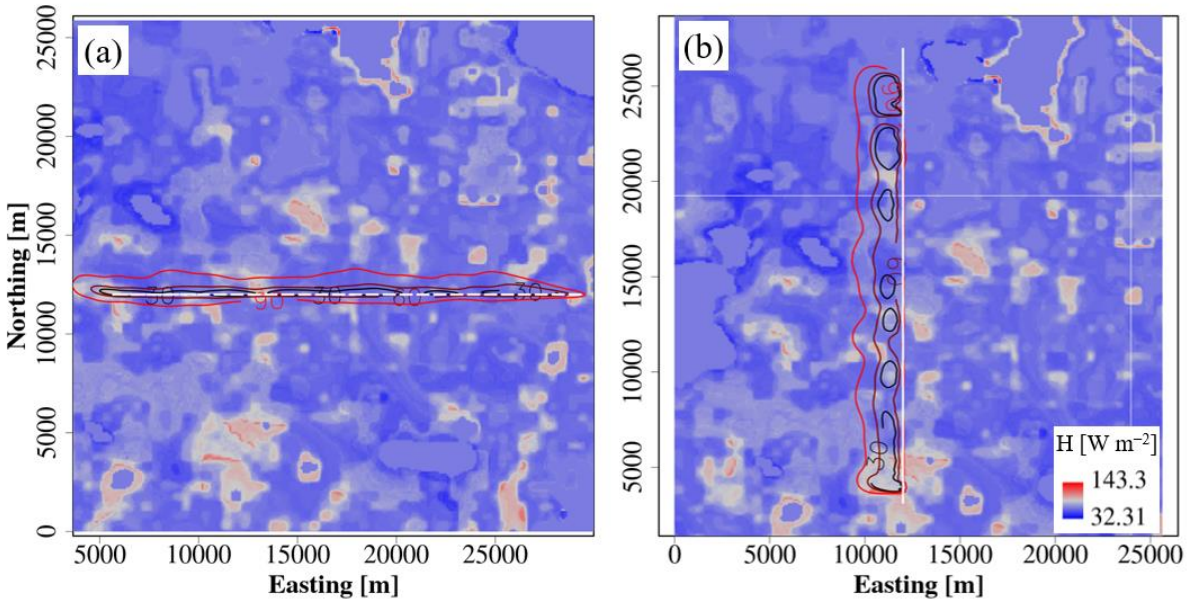


Figure 12. Example virtual flight tracks and their 30%, 60% and 90% source area contours superimposed over the LES sensible heat flux surface forcing (W m^{-2}). (a) Wind-parallel flights sample source areas that are elongated along the flight track leading to considerable overlap. (b) Wind-perpendicular flights sample less overlapping source areas along the flight track, and thus capture more independent samples and spatial variability.

Next, we address the design hypothesis H2 whether it is more informative to fly a finely-spaced pattern once, or to fly a coarsely-spaced pattern multiple times. Table 2 shows that the spatial coverage and energy balance ratio criteria are not particularly sensitive to the flight sequence. One exception is the particularly high and consistent improvement in the spatial patterning performance criterion of the alternating sequence (29.1±5.4%; Table 2 green cells). It is the only sequence that "carpets" the CHEESEHEAD19 domain side-to-side at fine time- and space increments. All other sequences progress in coarser increments, such as first completing all

parallel flight tracks and then revisiting the interspersed diagonal flight tracks. In the context of the 2 km horizontally spaced parallel flight tracks, K. Xu et al. (2017) report a 411±88 m characteristic surface length scale of landscape elements in the CHEESEHEAD19 domain. The finer increments of the parallel-diagonal alternating sequence let ERF relate drivers and responses closer to the characteristic surface length scale, and thus to reproduce the spatial patterning.

Table 2. Percent improvement of the joint tower and aircraft EC alternative OSDs relative to the tower-EC-only baseline OSD, aggregated over all four aircraft track angles on the mean wind direction. The results are shown as a function of optimality criterion (rows) and aircraft flying the grid pattern in various sequences (columns). The green cell highlights a marked improvement that is further discussed in the text.

Optimality criterion	**All sequences**	**Alternating**	**Outbound**	**Return**
Spatial coverage	23.6±2.2	23.0±3.0	23.5±1.8	24.4±1.9
Energy balance ratio	6.8±5.3	7.9±5.3	6.0±6.6	6.6±5.4
Spatial patterning	23.2±11.7	29.1±5.4	14.9±15.7	25.6±9.0

To summarize, flight patterns with a track angle in the range of 90°±45° on the mean wind direction yielded approximately double the performance improvement of wind-parallel patterns. This finding is irrespective of the flight sequence, but most consistent for the alternating flight sequence (21.6±11.5% – 22.6±9.4%; Table 3 green cells). Compared to the worst-case combination of wind-parallel flight patterns with the outbound flight sequence (9.6±11.1% improvement) this equates to doubling the scientific return.

Table 3. Percent improvement of the joint tower and aircraft EC alternative OSDs relative to the tower-EC-only baseline OSD, aggregated into a single score over all optimality criteria. The results are shown as a function of aircraft flying the grid pattern in various sequences (rows) and track angles on the mean wind direction (columns). Green cells highlight marked improvements that are further discussed in the text.

Flight sequence	**All angles**	**0°**	**45°**	**60°**	**90°**
All sequences	17.9±10.8	11.6±8.8	19.9±12.2	16.9±10.0	18.5±9.6
Alternating	20.0±10.2	13.3±11.4	22.6±9.4	21.6±11.5	22.4±11.7
Outbound	14.8±11.6	9.6±11.1	21.3±15.9	13.8±11.2	14.5±11.9
Return	18.9±10.7	13.4±10.3	21.8±17.7	20.5±10.2	19.8±6.2

3.4 Field campaign resources

Flying the grid pattern in the alternating sequence provided the best and most consistent results, while also being the most straightforward sequence for operational implementation. Further, to satisfy the 90°±45° track angle on mean wind condition we derived three rotationally symmetric sets of waypoints at 60° increments (Figure 13). Flying the numbered waypoint in ascending order results in three alternating flight sequences SE1, SW1 and WE1 with 330°, 30°, and 90° azimuth of the parallel tracks, respectively. Owing to rotational symmetry, flying the numbered waypoints in descending order results in three additional alternating flight sequences SE2, SW2 and WE2 with 150°, 210°, and 270° azimuth of the parallel tracks, respectively. Reversing the waypoint order allows the aircraft to progress through the flight tracks in downwind order for any given mean wind direction. This aims to reduce the space-time ambiguity resulting from airborne EC observing different surface conditions over hundreds of kilometers while at the same time the diurnal cycle progresses over the course of several hours. Lastly, to avoid the town and airfield of Park Falls, WI immediately west of the CHEESEHEAD19 domain, we shifted the WE1/WE2 set of waypoints 5 km to the east (Figure 13c).

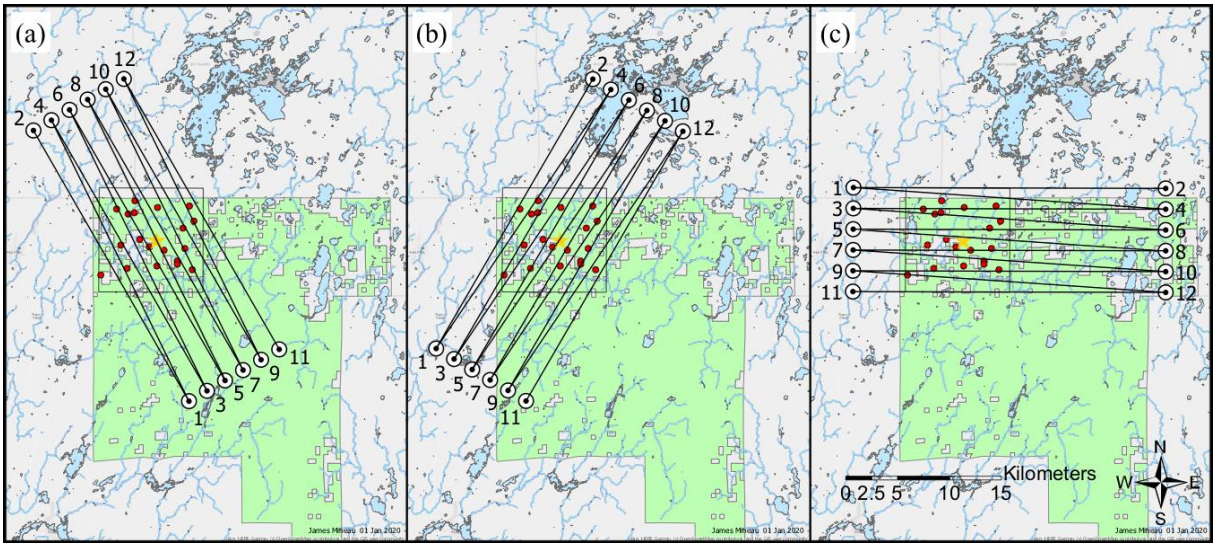


Figure 13. Three sets of waypoints define three distinct flight patterns, named after the relative location of their first two waypoints: (a) south-west (SW), (b) south-east (SE), and (c) west-east (WE). Flying the numbered waypoints either in ascending order (SW1, SE1, WE1) or descending order (SW2, SE2, WE2) resulted in six distinct flight sequences that maximize data coverage under different wind conditions. Map credit: James Mineau, University of Wisconsin – Madison.

To support daily flight planning we distilled the six alternating flight sequences into a flight planning wind rose (Figure 14). There we implemented the track angle condition by superimposing over a wind rose the wind sector aligned 90°±45° to the parallel tracks of each of the six alternating flight sequences. This allows determining the appropriate flight sequence as a function of the forecasted mean wind direction. For example, if experiencing southerlies (180°) the most suitable flight sequence is WE2. Owing to rotational symmetry, the wind sector for each flight sequence overlaps with each of its two neighbors by 30°. This provides a margin for accommodating changing synoptic conditions. For example, if experiencing south-southwesterlies (210°) in the morning the WE2 and SE1 flight sequences would be equally suitable. If however the mean direction is forecasted to shift to westerlies (270°) in the course of the day the SE1 flight sequence simplifies flight operation by satisfying the 90°±45° track angle on mean wind condition with a single flight sequence for a given day.

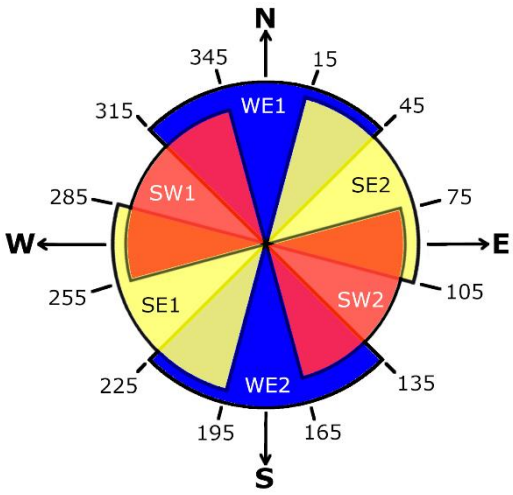


Figure 14. Flight planning wind rose to choose the appropriate flight sequence as a function of the forecasted mean wind direction. Owing to rotational symmetry, the wind sector for each flight sequence overlaps with each of its two neighbors by 30°. This provides a margin for accommodating changing synoptic conditions.

4 Discussion

Upon deriving the LES-ERF framework, we identified an optimal OSD that promises to more than double CHEESEHEAD19 scientific return. Here we initially discuss how these numerical gains relate to improving our ability for addressing CHEESEHEAD19 science objectives, and their limitations. We then examine how the resulting field campaign resources improved flight operation and crew safety by an order-of-magnitude. Lastly, we reflect our findings in the light of existing design approaches, provide general recommendations for future OSDs, and discuss remaining challenges and future work.

4.1 Optimizing the CHEESEHEAD19 observing system design

LES-ERF used three optimality criteria (Sect. 2.2; CR1 spatial coverage, CR2 energy balance ratio, CR3 spatial patterning) that we tailored to represent CHEESEHEAD19's science

741 objectives numerically. Furthermore, we identified two specific design hypotheses that we
742 postulate the science objectives, and hence optimality criteria to be sensitive to (Sect. 2.3; H1
743 track angle on mean wind, H2 fine vs. coarse flight sequence). CHEESEHEAD19's first science
744 objective O1 is to show that higher surface heterogeneity promotes energy transport in
745 atmospheric mesoscale eddies. Our ability to address this science objective increases with the
746 truthful reproduction of CR1 surface flux spatial coverage and CR3 spatial patterning. LES-ERF
747 allowed us to assess changes in these criteria resulting from the different OSDs by comparing
748 ERF flux map reproductions to the original LES surface flux forcing. We found that CR1 spatial
749 coverage is largely insensitive to either H1 track angle on mean wind (Table 1), and H2 fine vs.
750 coarse flight sequence (Table 2). Conversely, CR3 spatial patterning proved to be highly
751 sensitive to H1 track angle on mean wind. Track angles in the range of 90°±45° on the mean
752 wind yielded double to triple improvements over wind-parallel flights (Table 1). Similarly, we
753 showed that CR3 spatial patterning is sensitive to H2 fine vs. coarse flight sequence (Table 2).
754 The finely spaced "Alternating" sequence yielded the highest and most consistent spatial
755 patterning improvements of about 50% over the other flight sequences.

756 CHEESEHEAD19's second science objective O2 aims to account for energy transport in
757 mesoscale eddies and determine the "true" surface energy balance to improve model
758 representation of sub-grid processes. As such, our ability to address this science objective hinges
759 on improved closure of CR2 the energy balance ratio, which proved to be highly sensitive to H1
760 track angle on mean wind. Here, track angles in the range of 90°±45° on the mean wind yielded
761 quadruple to octuple improvements over wind-parallel flights (Table 1). On the other hand, CR2
762 energy balance ratio was comparatively insensitive to H2 fine vs. coarse flight sequence
763 (Table 2).

764 Finally, CHEESEHEAD19's third science objective O3 aims to demonstrate that ERF yields
765 representative fluxes at model grid scale regardless of mesoscale eddies. ERF's ability to
766 reproduce the surface flux is thus directly related to the combination of all criteria discussed
767 above, CR1 spatial coverage, CR2 energy balance ratio, and CR3 spatial patterning. From
768 aggregating over all optimality criteria into a single score, we demonstrated that overall
769 improvement is highly sensitive to H1 track angle on mean wind. Flight patterns with a track

angle in the range of 90°±45° on the mean wind yielded approximately double the performance improvement of wind-parallel patterns (Table 3). This finding is less sensitive to H2 fine vs. coarse flight sequence, though most consistent for the finely spaced "Alternating" sequence. Overall, this combination doubles the scientific return compared to the worst-case combination of wind-parallel flight patterns with the outbound flight sequence. Overall, the study hypothesis that CHEESEHEAD19 scientific return is sensitive to LES-ERF optimization can thus be accepted. On the other hand, the design hypothesis H1 that it is critical for airborne EC to measure perpendicular to the prevailing wind should be rejected, as up to ±45° tolerance yielded comparable results for CHEESEHEAD19 science objectives. Lastly, the design hypothesis H2 that it is more informative to fly a finely spaced pattern should be accepted, with most consistent results for the finely spaced "Alternating" sequence.

The field measurement resources (Sect. 3.4) encapsulate these findings into only three sets of waypoints and six alternating flight sequences incremented at 60° azimuth. These provide a balance between scientific fidelity and flight crew safety. On the one hand, the small number of waypoints and flight sequences is sufficient to maximize CHEESEHEAD19 scientific return by enabling to observe the 90°±45° track angle on mean wind condition at all times. Furthermore, the 60° incrementation in combination with the ±45° tolerance on perpendicularity to the mean wind provides 30° overlap among flight sequences to support decision-making during non-stationary synoptic conditions. On the other hand, the parsimonious number of only 6 flight sequences and an even smaller number of 3 sets of waypoints simplify flight planning and navigation. In combination with entirely avoiding the town and airfield of Park Falls this promotes flight crew safety by an order of magnitude compared to the originally envisioned 48 flight sequences. Specifically, it frees up the flight crew from arduous navigation details, thus reducing fatigue, increasing awareness during the 100 m low-level flight maneuvers, and ultimately reducing the margin for human error. Due to its vicinity to the 400 m tall tower and related in-flight safety concerns the central diagonal flight track in this study was not performed during the actual CHEESEHEAD19 field measurement campaign.

After deriving the above strategy, we detected an inconsistency in the vertical humidity profile that we used to initialize the LES and thus to produce the LES-ERF virtual observations.

Specifically, we had erroneously added a positive vertical humidity gradient at 350 m above ground instead of the negative vertical humidity gradient typically observed by radiosondes adjacent to the CHEESEHEAD19 domain (sign reversal). As a result, we detected an ABL height of only 500 m in the LES while field observations typically indicate >1 km during similar conditions around the CHEESEHEAD19 domain. In addition, during 1100 CST we detected a small entrainment flux of humid air from above the LES ABL into the dryer air below, which is not typical of summertime convective ABL conditions around the CHEESEHEAD19 domain. In the subsequent stages of LES-ERF, we used virtual tower EC observations at 49 m (N=20 towers), 112 m (N=1 tower), and 371 m (N=1 tower) above ground, and virtual airborne EC observations at 100 m above ground. At any given time, the surface fluxes prescribed in the LES were orders of magnitude larger compared to the entrainment fluxes. Hence, the surface fluxes dominated all virtual tower and airborne observations, possibly with exception of the 371 m tower that however still reported an average upward latent heat flux of 2.2±6.6 W m^{-2}. While the uncharacteristically shallow ABL height results in two to three times more pronounced vertical flux divergence, LES-ERF accounts for this by utilizing the relative measurement height in the ABL as an ERF driver. Furthermore, the study design cancels possible residual impacts on the CR2 energy balance ratio by normalizing all results for the alternative OSDs (tower and aircraft) with the results for the baseline OSD (tower-only). To summarize, the erroneous vertical humidity gradient resulted in a modelled LES atmosphere that was less specific to the CHEESEHEAD19 domain than originally envisioned. However, this should have little to no bearing on the general findings that informed the CHEESEHEAD19 OSD, owing to ERF accounting for vertical flux divergence and the normalized study design. If at all, surface heterogeneity scales across the CHEESEHEAD19 domain are more realistically reproduced compared to the idealized LES runs in many previous studies (e.g., Kanda et al., 2004; Sühring et al., 2018; K. Xu et al., 2020).

Overall, the application of the LES-ERF-derived field measurement resources enabled the successful acquisition of 14,400 km airborne data by the UWKA aircraft (Paleri et al., 2019). The 24 UWKA research flights and their on-site planning covered 480 flight legs during 72 hours of flight time and three CHEESEHEAD19 intensive observation periods (Butterworth et al., 2020). This further demonstrates the successful application of LES-ERF and its utility for

determining concise and adaptive field measurement resources that optimize the effectiveness and safety of research flights.

4.2 Benefits for coordinated environmental observations

LES-ERF extends on previous approaches to designing large-scale field campaigns such as FIFE, BOREAS, NOPEX, LITFASS-98, LITFASS-2003, MAGIM and ScaleX (Sect. 1). Specifically, LES-ERF not only utilizes but fully contextualizes expert knowledge by conducting virtual pre-field measurements in LES, and using evidence-driven ERF to quantify the resulting information contents.

For decades, LES "data from knowledge" studies have investigated surface-atmosphere interactions including energy balance processes (J. W. Deardorff, 1972; Etling & Brown, 1993; Kanda et al., 2004; Sühring & Raasch, 2013; Wyngaard & Brost, 1984). Indeed, LES have become useful to contextualize observational phenomena with increasingly complex feedbacks, including natural resources such as air quality (Khan et al., 2020; Y. Zhang et al., 2020). However, the resulting data is detailed to a point where it becomes challenging to fully utilize the provided information for extracting and describing the phenomena of interest. K. Xu et al. (2020) point to a possible solution to this dilemma, by complementing detailed LES outputs with the dedicated ERF "knowledge from data" approach. Here, we took a next step and demonstrated the usefulness of the LES-ERF symbiosis for designing real-world field measurements. Using LES for OSD has been a rare application to date (Cortina & Calaf, 2017; Gehrke et al., 2019), and to our knowledge the present study is the first of its kind that empowers investigators to harness the combined power of complementing NSs and data mining for this purpose.

OSSEs are widely used in the earth-system sciences to predict the performance of major, long-term research equipment and facility investments (Hargrove & Hoffman, 2004; Hoffman & Atlas, 2016; Lucas et al., 2015; Masutani et al., 2010; J. Park & Kim, 2020; L. Zhang & Pu, 2010). The LES-ERF symbiosis now provides the necessary resolution of time, space, and processes to make the power of OSSEs also accessible for designing field measurements at smaller and previously inaccessible scales. Specifically, the CHEESEHEAD19 example application at the interface of meso- and microscale meteorology demonstrated a new degree of

realism and explicitness in maximizing the joint information from ground-based, airborne, and spaceborne observations for scaling and modeling.

Building on this central property, LES-ERF is modularly extensible in multiple directions. For example, LES-ERF can integrate new types of observations in addition to tower, aircraft, and satellite observations, so long as their source areas are readily quantifiable. Examples are remote sensing of the atmosphere (Helbig et al., 2020; Wulfmeyer et al., 2018) and soil and biometric observations (Metzger, Ayres, Durden, et al., 2019). This provides a promising avenue to maximize cross-disciplinary, cross-project, and ultimately cross-institutional synergies. For example, such as through simulating the design of super-sites that envision to synergize diverse observational infrastructures including from the US National Science Foundation's National Center for Atmospheric Research and National Ecological Observatory Network (Metzger, Ayres, Desai, et al., 2019). Then upon completion of the planned field measurements, the real-world data can immediately substitute the LES "data from knowledge" module, while the ERF "knowledge from data" module continues to perform the intended end-to-end analyses. LES-ERF thus provides a framework to prepare and test the quantification of science objectives well ahead of the actual field measurements, thus reducing latency from field data capture to knowledge creation. Ultimately, the characteristics of LES-ERF make the power of OSSEs accessible to an entirely new range of use cases. Examples include natural climate solutions (Hemes et al., in review), emission inventory validation (Desjardins et al., 2018), urban air quality (Squires et al., 2020), industry leak detection (Kohnert et al., 2017), and multi-species applications (Vaughan et al., 2017).

4.3 Remaining challenges and future work

Notwithstanding these key benefits, a LES-ERF study such as presented here adds labor, computing, and hence funding requirements ahead of the actual field measurements. Considering a typical research grant cycle, one would ideally perform the LES-ERF OSDs prior to submitting a funding proposal, or at least request some level of design flexibility. We conducted the present study over the course of approximately three months, und utilized the labor and computing resources summarized in Table 4. Overall, we spent ~480 h of labor, or three person-months, of which the LES and ERF analyses consumed ~40% and ~60%, respectively. The main labor

886 drivers were study conceptualization and setup including data acquisition for LES boundary
887 conditions. It is possible to perform these steps well in advance, e.g. to reduce the LES-ERF
888 effort between grant receipt and field measurements, which is typically also a period with high-
889 demand for overall coordination. The 50 h spent on ERF interpretation also included active
890 dialogue and iteration with the flight crew, resulting in balanced resources for airborne operation.

891 Table 4 also shows how LES and ERF differed in their computational needs. LES demanded
892 high-performance computing with 230,000 CPU hours and up to 120 TB memory, which we
893 primarily performed on the US National Center for Atmospheric Research Cheyenne
894 supercomputer (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cheyenne).
895 Conversely, ERF required a high-throughput computing architecture, for which we primarily
896 used the US National Science Foundation's CyVerse open science workspace
897 (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cyverse). Overall, the strong

898

899 Table 4. Labor and computing resources utilized for deriving the CHEESEHEAD19 observing
900 system design, separately for Large Eddy Simulations (LES) and Environmental Response
901 Functions (ERF).

Resource	**LES**	**ERF**
Total Labor	180 h	300 h
Conceptualization	30 h	70 h
Setup	110 h	130 h
Analysis	20 h	50 h
Interpretation	20 h	50 h
Computing architecture	High-performance	High-throughput
CPU hours	230,000	7,000
CPUs	600 – 1,800	2 – 16
Memory	2 – 120 TB	16 – 128 GB
Data produced	210 GB	4 GB

902

data requirements of ERF, including use of high-frequency EC data, currently drive LES-ERF computational needs. Investigations into relaxing ERF data requirements while retaining overall performance are in progress. This could permit generating the necessary virtual observations with NSs that substantially reduce resource demand compared to LES, such as Reynolds-averaged Navier-Stokes (Santiago et al., 2010; Sogachev et al., 2002). In turn, this could enable application of a generalized "NS-ERF" approach to OSSEs for use cases that require consideration of more extensive space-, time-, and disciplinary domains. Furthermore, a unified graphical user interface could aid accessibility and usability to better support investigators from diverse backgrounds.

In Sect. 4.1 we discussed several sources of uncertainty that emanated from the specific numerical analyses chosen to optimize the CHEESEHEAD19 flight strategy. In addition, sources of uncertainty that surround the LES-ERF concept as a whole also warrant discussion. One of the strengths of OSSEs in general and LES-ERF in particular is to quantify the efficacy of candidate OSDs for cross-disciplinary applications. However, individual disciplines themselves often invoke very specific criteria and assumptions so their contributions to the overall project are valid (Sect. 1). Also determining the OSD trade-offs for meeting these discipline-specific criteria requirements could complement LES-ERF to an end-to-end science traceability assessment. One direction of future work could use the CHEESEHEAD19 field measurements to derive and evaluate such an end-to-end assessment in general and the presented OSD results in particular.

Furthermore, earth system observations are highly variable in their space-time extent and resolution (Figure 1). However, data overlays such as done in ERF require a "least common denominator" space-time resolution among all considered observations. Attaining this least common denominator while retaining quasi-continuous data coverage remains an observational challenge, even for WMO Essential Climate Variables such as land surface temperature. Toolkits that leverage multi-sensor data fusion to provide the necessary resolution and coverage to support plot- to landscape-scale research are only recently emerging (Pincebourde & Salle, 2020; P. Wu et al., 2013).

Earth system observations are also variable in their space-time representations. These include gridded remote sensing observations in Eulerian coordinates, and EC heat flux observations in

Lagrangian coordinates (Metzger, 2018). Data overlays among these observations leverage source area models, which connect the two representations in space and time (Leclerc & Foken, 2014). However, e.g.Bertoldi et al. (2013); K. Xu et al. (2020) point out a possible dependency of the source area attribution performance on the thermodynamic properties of the quantity observed in Lagrangian coordinates. Robust data overlays across coordinate representations might thus depend on separate source area considerations for neutral density vs. self-buoyant quantities.

5 Conclusions

Surface-atmosphere synthesis is traditionally in the vanguard of interdisciplinary research, with efforts ranging from empirical studies over theoretical generalizations to NSs. More recently, data-intensive information discovery promises to further expand our insight into momentum, energy, water, and trace gas cycling. However, "data deluge" rather than the next interdisciplinary breakthrough can result from poor information overlap among ground, airborne and satellite observations, as well as numerical models. Scientific return hinges on our ability to merge information among these perspectives, for which the pre-field stage provides a unique opportunity to optimize the study design accordingly.

We harnessed this opportunity by catalyzing recent advances in conducting virtual experiments within high-resolution LESs ("data from knowledge") and physics-guided data science ("knowledge from data") to create the LES-ERF approach. Traditional data capture focusses on intra-disciplinary best practices, and the cross-discipline explanatory power only becomes clear afterward. In contrast, LES-ERF explores tolerances ("value engineering") in a numerical framework ahead of the actual field deployments, which offers a wide margin for improving cross-discipline post-field synthesis. For the CHEESEHEAD19 OSD, we used LES-ERF to maximize the information overlap across micro- and mesometeorological space and time scales. To date, these scales have predominantly been dealt with in a discontinuous fashion, which we overcame by combining cross-platform flux tower and flux aircraft responses in a single ERF for the first time. This demonstrated that a carefully designed flight strategy has the potential to double CHEESEHEAD19 scientific return across two specific design hypotheses, and to order-of-magnitude improve flight operation and crew safety.

LES-ERF thus makes the benefits of OSSEs accessible for maximizing the return of cross-disciplinary field measurements that previously had to rely on experience and expert knowledge alone. This property transcends academic field measurements such as presented here, and can inform natural climate solutions, emission inventory validation, urban air quality, industry leak detection and multi-species applications, among other use cases.

Data Availability Statement

All data used in this study are available per CHEESEHEAD Data Policy (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/data-policy) from the FAIR-compliant CyVerse Data Commons (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/data-commons). The top-level document "readme_Metzger-et-al_2021_OSSE-LES-ERF.md" provides specific dataset locations for individual processing steps. For final manuscript publication, we plan to curate these data into the FAIR-compliant CHEESEHEAD Data Archive Center at the National Center for Atmospheric Research, Earth Observing Laboratory (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/data-repo), mint a DOI and cite the dataset as full reference in the reference list. Software used in this study are available per CHEESEHEAD Code Policy (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/code-policy), and are either already available from or being developed into public code repositories.

Acknowledgements

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 and AGS-1822420]. 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. We thank University of Wyoming King Air pilot Edward Sigel and flight crew for their flight strategy input, and James Mineau for creating the flight leg maps. L. Wanner's and M. Mauder's contributions were supported by Deutsche Forschungsgemeinschaft (DFG) Grant #406980118.

References

Atlas, R., Hoffman, R. N., Ma, Z., Emmitt, G. D., Wood, S. A., Jr., Greco, S., et al. (2015). Observing System Simulation Experiments (OSSEs) to evaluate the potential impact of an Optical Autocovariance Wind Lidar (OAWL) on numerical weather prediction. *Journal of Atmospheric and Oceanic Technology, 32*(9), 1593-1613. https://doi.org/10.1175/JTECH-D-15-0038.1

Avissar, R., & Schmidt, T. (1998). An evaluation of the scale at which ground-surface heat flux patchiness affects the convective boundary layer using large-eddy simulations. *Journal of the Atmospheric Sciences, 55*(16), 2666-2689. https://doi.org/10.1175/1520-0469(1998)055<2666:AEOTSA>2.0.CO;2

Bertoldi, G., Kustas, W., & Albertson, J. (2013). Evaluating source area contributions from aircraft flux measurements over heterogeneous land using large-eddy simulation. *Boundary-Layer Meteorology, 147*(2), 261-279. https://doi.org/10.1007/s10546-012-9781-y

Beyrich, F., Herzog, H. J., & Neisser, J. (2002). The LITFASS project of DWD and the LITFASS-98 experiment: The project strategy and the experimental setup. *Theoretical and Applied Climatology, 73*(1), 3-18. https://doi.org/10.1007/s00704-002-0690-8

Beyrich, F., & Mengelkamp, H.-T. (2006). Evaporation over a heterogeneous land surface: EVA_GRIPS and the LITFASS-2003 experiment - an overview. *Boundary-Layer Meteorology, 121*(1), 5-32. https://doi.org/10.1007/s10546-006-9079-z

Brown, J. R., & Fehige, Y. (2019). Thought experiments. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 ed., pp. https://plato.stanford.edu/archives/win2019/entries/thought-experiment/). Stanford, USA: Metaphysics Research Lab, Stanford University.

Butterbach-Bahl, K., Kögel-Knabner, I., & Han, X. (2011). Steppe ecosystems and climate and land-use changes—vulnerability, feedbacks and possibilities for adaptation. *Plant and Soil, 340*(1), 1-6. https://doi.org/10.1007/s11104-010-0651-4

1016 Butterworth, B. J., Desai, A. R., Metzger, S., Townsend, P. A., Schwartz, M. D., Petty, G. W., et
1017 al. (2020). Connecting land-atmosphere interactions to surface heterogeneity in
1018 CHEESEHEAD19. *Bulletin of the American Meteorological Society*, 1-71.
1019 https://doi.org/10.1175/BAMS-D-19-0346.1

1020 Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-P., & Foken, T. (2014). Extension of the
1021 averaging time in eddy-covariance measurements and its effect on the energy balance closure.
1022 *Boundary-Layer Meteorology, 152*(3), 303-327. https://doi.org/10.1007/s10546-014-9922-6

1023 Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., et al. (2011).
1024 Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research
1025 Network based on remote sensing and footprint modelling. *Agricultural and Forest Meteorology,*
1026 *151*(1), 87-100. https://doi.org/10.1016/j.agrformet.2010.09.005

1027 Chu, H., Baldocchi, D. D., John, R., Wolf, S., & Reichstein, M. (2017). Fluxes all of the time? A
1028 primer on the temporal representativeness of FLUXNET. *Journal of Geophysical Research:*
1029 *Biogeosciences, 122*(2), 289-307. https://doi.org/10.1002/2016JG003576

1030 Cortina, G., & Calaf, M. (2017). Turbulence upstream of wind turbines: A large-eddy simulation
1031 approach to investigate the use of wind lidars. *Renewable Energy, 105*, 354-365.
1032 https://doi.org/10.1016/j.renene.2016.12.069

1033 Deardorff, J. W. (1972). Numerical investigation of neutral and unstable planetary boundary
1034 layers. *Journal of the Atmospheric Sciences, 29*(1), 91-115. https://doi.org/10.1175/1520-
1035 0469(1972)029<0091:NIONAU>2.0.CO;2

1036 Deardorff, J. W. (1980). Stratocumulus-capped mixed layers derived from a three-dimensional
1037 model. *Boundary-Layer Meteorology, 18*(4), 495-527. https://doi.org/10.1007/BF00119502

1038 Desai, A. (1996 -). AmeriFlux US-PFa Park Falls/WLEF, Dataset.

1039 Desai, A. R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., et al. (2015).
1040 Landscape-level terrestrial methane flux observed from a very tall tower. *Agricultural and*
1041 *Forest Meteorology, 201*(0), 61-75. https://doi.org/10.1016/j.agrformet.2014.10.017

1042 Desjardins, R. L., Worth, D. E., Pattey, E., VanderZaag, A., Srinivasan, R., Mauder, M., et al.
1043 (2018). The challenge of reconciling bottom-up agricultural methane emissions inventories with
1044 top-down measurements. *Agricultural and Forest Meteorology, 248*, 48-59.
1045 https://doi.org/10.1016/j.agrformet.2017.09.003

1046 Eddy, A. (1974). An approach to the design of meteorological field experiments. *Monthly*
1047 *Weather Review, 102*(10), 702-707. https://doi.org/10.1175/1520-
1048 0493(1974)102<0702:AATTDO>2.0.CO;2

1049 Etling, D., & Brown, R. A. (1993). Roll vortices in the planetary boundary layer: A review.
1050 *Boundary-Layer Meteorology, 65*(3), 215-248. https://doi.org/10.1007/BF00705527

1051 Foken, T. (2017). *Micrometeorology* (2 ed.). Berlin, Heidelberg: Springer.

1052 Foken, T., Aubinet, M., Finnigan, J. J., Leclerc, M. Y., Mauder, M., & Paw U, K. T. (2011).
1053 Results of a panel discussion about the energy balance closure correction for trace gases. *Bulletin*
1054 *of the American Meteorological Society, 92*(4), ES13-ES18.
1055 https://doi.org/10.1175/2011BAMS3130.1

1056 Gao, Z., Liu, H., Russell, E. S., Huang, J., Foken, T., & Oncley, S. P. (2016). Large eddies
1057 modulating flux convergence and divergence in a disturbed unstable atmospheric surface layer.
1058 *Journal of Geophysical Research: Atmospheres, 121*(4), 1475-1492.
1059 https://doi.org/10.1002/2015jd024529

1060 Gehrke, K., Böske, L., & Beyrich, F. (2019). *Analyse verschiedener Doppler-Lidar Scan-*
1061 *Strategien zur Ableitung des mittleren Windvektors in der konvektiven Grenzschicht auf Basis*
1062 *von Large-Eddy-Simulationen*. Paper presented at the DACH conference of German, Austrian
1063 and Swiss meteorologists, Garmisch-Partenkirchen, Germany.
1064 https://meetingorganizer.copernicus.org/DACH2019/DACH2019-241-1.pdf

1065 Halldin, S., Gryning, S. E., Gottschalk, L., Jochum, A., Lundin, L. C., & Van de Griend, A. A.
1066 (1999). Energy, water and carbon exchange in a boreal forest landscape - NOPEX experiences.
1067 *Agricultural and Forest Meteorology, 98-99*, 5-29. https://doi.org/10.1016/S0168-
1068 1923(99)00148-3

Hargrove, W. W., & Hoffman, F. M. (2004). Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management, 34*(1), S39-S60. https://doi.org/10.1007/s00267-003-1084-0

Helbig, M., Gerken, T., Beamesderfer, E., Baldocchi, D. D., Banerjee, T., Biraud, S. C., et al. (2020). *White Paper: Understanding land-atmosphere interactions through tower-based flux and continuous atmospheric boundary layer measurements*. Retrieved from Berkeley, CA, U.S.A.: https://ameriflux.lbl.gov/community/highlight/whitepaper-understanding-land-atmosphere-interactions-through-tower-based-flux-and-continuous-atmospheric-boundary-layer-measurements/

Hemes, K. S., Runkle, B., Novick, K., Baldocchi, D. D., & Field, C. B. (in review). The role of ecosystem-scale flux measurements in assessing implementation of natural climate solutions.

Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, USA: Microsoft Research.

Hoffman, R. N., & Atlas, R. (2016). Future observing system simulation experiments. *Bulletin of the American Meteorological Society, 97*(9), 1601-1616. https://doi.org/10.1175/BAMS-D-15-00200.1

Kaharabata, S. K., Schuepp, P. H., Ogunjemiyo, S., Shen, S., Leclerc, M. Y., Desjardins, R. L., & MacPherson, J. I. (1997). Footprint considerations in BOREAS. *Journal of Geophysical Research, [Atmospheres], 102*(D24), 29113-29124. https://doi.org/10.1029/97JD02559

Kaminski, T., Rayner, P. J., Voßbeck, M., Scholze, M., & Koffi, E. (2012). Observing the continental-scale carbon balance: Assessment of sampling complementarity and redundancy in a terrestrial assimilation system by means of quantitative network design. *Atmospheric Chemistry and Physics, 12*(16), 7867-7879. https://doi.org/10.5194/acp-12-7867-2012

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. https://doi.org/10.1023/B:BOUN.0000007225.45548.7a

1095 Keller, M., Schimel, D. S., Hargrove, W. W., & Hoffman, F. M. (2008). A continental strategy
1096 for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment,*
1097 *6*(5), 282-284. https://doi.org/10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2

1098 Khan, B., Banzhaf, S., Chan, E. C., Forkel, R., Kanani-Sühring, F., Ketelsen, K., et al. (2020).
1099 Development of an atmospheric chemistry model coupled to the PALM model system 6.0:
1100 Implementation and first applications. *Geosci. Model Dev. Discuss., 2020*, 1-34.
1101 https://doi.org/10.5194/gmd-2020-286

1102 Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2004). A simple parameterisation for
1103 flux footprint predictions. *Boundary-Layer Meteorology, 112*(3), 503-523.
1104 https://doi.org/10.1023/B:BOUN.0000030653.71031.96

1105 Koffi, E. N., Rayner, P. J., Scholze, M., Chevallier, F., & Kaminski, T. (2013). Quantifying the
1106 constraint of biospheric process parameters by CO_2 concentration and flux measurement
1107 networks through a carbon cycle data assimilation system. *Atmospheric Chemistry and Physics,*
1108 *13*(21), 10555-10572. https://doi.org/10.5194/acp-13-10555-2013

1109 Kohnert, K., Serafimovich, A., Metzger, S., Hartmann, J., & Sachs, T. (2017). Strong geologic
1110 methane emissions from discontinuous terrestrial permafrost in the Mackenzie Delta, Canada.
1111 *Scientific Reports, 7*(1), 5828. https://doi.org/10.1038/s41598-017-05783-2

1112 Kumar, J., Hoffman, F. M., Hargrove, W. W., & Collier, N. (2016). Understanding the
1113 representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements.
1114 *Earth Syst. Sci. Data Discuss., 2016*, 1-25. https://doi.org/10.5194/essd-2016-36

1115 Lauvaux, T., Schuh, A. E., Bocquet, M., Wu, L., Richardson, S., Miles, N., & Davis, K. J.
1116 (2012). Network design for mesoscale inversions of CO_2 sources and sinks. *Tellus B: Chemical*
1117 *and Physical Meteorology, 64*(1), 17980. https://doi.org/10.3402/tellusb.v64i0.17980

1118 Leclerc, M. Y., & Foken, T. (2014). *Footprints in micrometeorology and ecology* (1st ed.).
1119 Berlin, Heidelberg, Germany: Springer.

Lenschow, D. H., 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. https://doi.org/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. https://doi.org/10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2

Loescher, H., Ayres, E., Duffy, P., Luo, H., & Brunke, M. (2014). Spatial variation in soil properties among North American ecosystems and guidelines for sampling designs. *PLoS One, 9*(1), e83216. https://doi.org/10.1371/journal.pone.0083216

Lucas, D. D., Yver Kwok, C., Cameron-Smith, P., Graven, H., Bergmann, D., Guilderson, T. P., et al. (2015). Designing optimal greenhouse gas observing networks that consider performance and cost. *Geosci. Instrum. Method. Data Syst., 4*(1), 121-137. https://doi.org/10.5194/gi-4-121-2015

Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., et al. (2017). Detecting impacts of extreme events with ecological in situ monitoring networks. *Biogeosciences, 14*(18), 4255-4277. https://doi.org/10.5194/bg-14-4255-2017

Mahrt, L. (1996). The bulk aerodynamic formulation over heterogeneous surfaces. *Boundary-Layer Meteorology, 78*(1-2), 87-119. https://doi.org/10.1007/BF00122488

Margairaz, F., Pardyjak, E. R., & Calaf, M. (2020). Surface thermal heterogeneities and the atmospheric boundary layer: The relevance of dispersive fluxes. *Boundary-Layer Meteorology, 175*(3), 369-395. https://doi.org/10.1007/s10546-020-00509-w

Maronga, B., Banzhaf, S., Burmeister, C., Esch, T., Forkel, R., Fröhlich, D., et al. (2020). Overview of the PALM model system 6.0. *Geosci. Model Dev., 13*(3), 1335-1372. https://doi.org/10.5194/gmd-13-1335-2020

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., et al. (2015). The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and

oceanic flows: model formulation, recent developments, and future perspectives. *Geosci. Model Dev., 8*(8), 2515-2551. https://doi.org/10.5194/gmd-8-2515-2015

Mason, P. J. (1988). The formation of areally-averaged roughness lengths. *Quarterly Journal of the Royal Meteorological Society, 114*(480), 399-420. https://doi.org/10.1002/qj.49711448007

Masutani, M., Woollen, J. S., Lord, S. J., Emmitt, G. D., Kleespies, T. J., Wood, S. A., et al. (2010). Observing system simulation experiments at the National Centers for Environmental Prediction. *Journal of Geophysical Research: Atmospheres, 115*(D7). https://doi.org/10.1029/2009JD012528

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], 112*(D13), D13112. https://doi.org/10.1029/2006JD008133

Mauder, M., Foken, T., & Cuxart, J. (2020). Surface-energy-balance closure over land: A review. *Boundary-Layer Meteorology*. https://doi.org/10.1007/s10546-020-00529-6

Mauder, M., Jegede, O. O., Okogbue, E. C., Wimmer, F., & Foken, T. (2007). Surface energy balance measurements at a tropical site in West Africa during the transition from dry to wet season. *Theoretical and Applied Climatology, 89*(3), 171-183. https://doi.org/10.1007/s00704-006-0252-6

Metzger, S. (2018). Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ observations. *Agricultural and Forest Meteorology, 255*, 68-80. https://doi.org/10.1016/j.agrformet.2017.08.037

Metzger, S., Ayres, E., Desai, A., Durden, D., Florian, C., Lee, R., et al. (2019). *Synthesized observations and processes for plot- to landscape-scale research, in: NCAR and NEON Town Hall TH13M*. Paper presented at the 52nd AGU annual Fall Meeting, San Francisco, U.S.A.

Metzger, S., Ayres, E., Durden, D., Florian, C., Lee, R., Lunch, C., et al. (2019). From NEON field sites to data portal: a community resource for surface–atmosphere research comes online.

Bulletin of the American Meteorological Society, 100(11), 2305-2325. https://doi.org/10.1175/bams-d-17-0307.1

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. *Atmospheric Measurement Techniques, 5*(7), 1699-1717. https://doi.org/10.5194/amt-5-1699-2012

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

Metzger, S., Xu, K., Desai, A. R., Taylor, J. R., Kljun, N., Schneider, D., et al. (2013). *Spatio-temporal rectification of tower-based eddy-covariance flux measurements for consistently informing process-based models*. Paper presented at the 46th AGU annual Fall Meeting, San Francisco, U.S.A.

Moeng, C.-H., & Rotunno, R. (1990). Vertical-velocity skewness in the buoyancy-driven boundary layer. *Journal of Atmospheric Sciences, 47*(9), 1149-1162. https://doi.org/10.1175/1520-0469(1990)047<1149:vvsitb>2.0.co;2

Montanari, R., Souza, G. S. A., Pereira, G. T., Marques, J., Siqueira, D. S., & Siqueira, G. M. (2012). The use of scaled semivariograms to plan soil sampling in sugarcane fields. *Precision Agriculture, 13*(5), 542-552. https://doi.org/10.1007/s11119-012-9265-6

Munger, J. W., Loescher, H. W., & Luo, H. Y. (2012). Measurement, tower, and site design considerations. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy covariance: A practical guide to measurement and data analysis* (pp. 21-58). Dordrecht, Heidelberg, London, New York: Springer.

Norton, J. (1991). Thought experiments in Einstein's work. In T. Horowitz & G. J. Massey (Eds.), *Thought Experiments in Science and Philosophy* (1st ed., pp. 335). Lanham, USA: Rowman & Littlefield.

Paleri, S., Desai, A. R., Metzger, S., Mauder, M., Plummer, D. M., & Wanner, L. (2019). *Mesoscale flux contributions to surface-atmosphere interactions across a heterogeneous mid-latitude landscape*. Paper presented at the 52nd AGU annual Fall Meeting, San Francisoco, U.S.A. https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/603722

Palmerino, C. R. (2018). Discussing what would happen: The role of thought experiments in Galileo's dialogues. *Philosophy of Science, 85*(5), 906-918. https://doi.org/10.1086/699717

Park, J., & Kim, H. M. (2020). Design and evaluation of CO_2 observation network to optimize surface CO_2 fluxes in Asia using observation system simulation experiments. *Atmospheric Chemistry and Physics, 20*(8), 5175-5195. https://doi.org/10.5194/acp-20-5175-2020

Park, R. (1998). *Value engineering: A plan for invention*. Boca Raton, USA: CRC Press.

Petty, G. W. (2020). Sampling error in aircraft flux measurements based on a high-resolution Large Eddy Simulation of the marine boundary layer. *Atmos. Meas. Tech. Discuss., 2020*, 1-33. https://doi.org/10.5194/amt-2020-235

Pincebourde, S., & Salle, A. (2020). On the importance of getting fine-scale temperature records near any surface. *Global Change Biology, 26*(11), 6025-6027. https://doi.org/10.1111/gcb.15210

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven earth system science. *Nature, 566*(7743), 195-204. https://doi.org/10.1038/s41586-019-0912-1

Santiago, J. L., Dejoan, A., Martilli, A., Martin, F., & Pinelli, A. (2010). Comparison between Large-Eddy Simulation and Reynolds-Averaged Navier-Stokes computations for the MUST field experiment. Part I: Study of the flow for an incident wind directed perpendicularly to the front array of containers. *Boundary-Layer Meteorology, 135*(1), 109-132. https://doi.org/10.1007/s10546-010-9466-3

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

1224 Schröter, M., Bange, J., & Raasch, S. (2000). Simulated airborne flux measurements in a LES
1225 generated convective boundary layer. *Boundary-Layer Meteorology*, *95*(3), 437-456.
1226 https://doi.org/10.1023/A:1002649322001

1227 Schuepp, P. H., Leclerc, M. Y., MacPherson, J. I., & Desjardins, R. L. (1990). Footprint
1228 prediction of scalar fluxes from analytical solutions of the diffusion equation. *Boundary-Layer*
1229 *Meteorology, 50*(1), 355-373. https://doi.org/10.1007/BF00120530

1230 Sellers, P., Hall, F., Margolis, H., Kelly, B., Baldocchi, D., den Hartog, G., et al. (1995). The
1231 Boreal Ecosystem-Atmosphere Study (BOREAS): An overview and early results from the 1994
1232 field year. *Bulletin of the American Meteorological Society*, *76*(9), 1549-1577.
1233 https://doi.org/10.1175/1520-0477(1995)076<1549:TBESAO>2.0.CO;2

1234 Sellers, P. J., Hall, F. G., Asrar, G., Strebel, D. E., & Murphy, R. E. (1988). The First ISLSCP
1235 Field Experiment (FIFE). *Bulletin of the American Meteorological Society*, *69*(1), 22-27.
1236 https://doi.org/10.1175/1520-0477(1988)069<0022:TFIFE>2.0.CO;2

1237 Serafimovich, A., Metzger, S., Hartmann, J., Kohnert, K., Zona, D., & Sachs, T. (2018).
1238 Upscaling surface energy fluxes over the North Slope of Alaska using airborne eddy-covariance
1239 measurements and environmental response functions. *Atmospheric Chemistry and Physics*,
1240 *18*(13), 10007-10023. https://doi.org/10.5194/acp-18-10007-2018

1241 Sogachev, A., Menzhulin, G. V., Heimann, M., & Lloyd, J. (2002). A simple three-dimensional
1242 canopy – planetary boundary layer simulation model for scalar concentrations and fluxes. *Tellus*
1243 *B: Chemical and Physical Meteorology*, *54*(5), 784-819.
1244 https://doi.org/10.3402/tellusb.v54i5.16729

1245 Squires, F. A., Nemitz, E., Langford, B., Wild, O., Drysdale, W. S., Acton, W. J. F., et al. (2020).
1246 Measurements of traffic-dominated pollutant emissions in a Chinese megacity. *Atmospheric*
1247 *Chemistry and Physics*, *20*(14), 8737-8761. https://doi.org/10.5194/acp-20-8737-2020

1248 Steinfeld, G., Letzel, M. O., Raasch, S., Kanda, M., & Inagaki, A. (2007). Spatial
1249 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. https://doi.org/10.1007/s10546-006-9133-x

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., et al. (2013). A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and Forest Meteorology, 171–172*(0), 137-152. https://doi.org/10.1016/j.agrformet.2012.11.004

Sühring, M., Metzger, S., Xu, K., Durden, D., & Desai, A. (2018). Trade-offs in flux disaggregation: a large-eddy simulation study. *Boundary-Layer Meteorology, 170*(1), 69-93. journal article. https://doi.org/10.1007/s10546-018-0387-x

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*, 1-23. https://doi.org/10.1007/s10546-013-9822-1

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, 116*, G00J04. https://doi.org/10.1029/2010JG001562

Taylor, G. I. (1915). Eddy motion in the atmosphere. *Philosophical Transactions of the Royal Society of London, A 215*, 1-26.

Tohidi, H. (2011). Review the benefits of using value engineering in information technology project management. *Procedia Computer Science, 3*, 917-924. https://doi.org/10.1016/j.procs.2010.12.150

Vaughan, A. R., Lee, J. D., Shaw, M. D., Misztal, P. K., Metzger, S., Vieno, M., et al. (2017). VOC emission rates over London and South East England obtained by airborne eddy covariance. *Faraday Discussions, 200*(0), 599-620. https://doi.org/10.1039/C7FD00002B

1274 Villarreal, S., Guevara, M., Alcaraz-Segura, D., & Vargas, R. (2019). Optimizing an
1275 environmental observatory network design using publicly available data. *Journal of Geophysical*
1276 *Research: Biogeosciences, 124*(7), 1812-1826. https://doi.org/10.1029/2018jg004714

1277 Wicker, L. J., & Skamarock, W. C. (2002). Time-Splitting Methods for Elastic Models Using
1278 Forward Time Schemes. *Monthly Weather Review, 130*(8), 2088-2097.
1279 https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2

1280 Wiens, R. C., Neugebauer, M., Reisenfeld, D. B., Moses, R. W., Nordholt, J. E., & Burnett, D. S.
1281 (2003). Genesis solar wind concentrator: Computer simulations of performance under solar wind
1282 conditions. *Space Science Reviews, 105*(3), 601-626. https://doi.org/10.1023/A:1024474028352

1283 Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., et al.
1284 (2009). Improving land surface models with FLUXNET data. *Biogeosciences, 6*(7), 1341-1359.
1285 https://doi.org/10.5194/bg-6-1341-2009

1286 Williamson, J. H. (1980). Low-storage Runge-Kutta schemes. *Journal of Computational Physics,*
1287 *35*(1), 48-56. https://doi.org/10.1016/0021-9991(80)90033-9

1288 Wolf, B., Chwala, C., Fersch, B., Garvelmann, J., Junkermann, W., Zeeman, M. J., et al. (2017).
1289 The SCALEX campaign: Scale-crossing land surface and boundary layer processes in the
1290 TERENO-preAlpine Observatory. *Bulletin of the American Meteorological Society, 98*(6), 1217-
1291 1234. https://doi.org/10.1175/BAMS-D-15-00277.1

1292 Wu, D., Wang, Z., Wechsler, P., Mahon, N., Deng, M., Glover, B., et al. (2016). Airborne
1293 compact rotational Raman lidar for temperature measurement. *Optics Express, 24*(18), A1210-
1294 A1223. https://doi.org/10.1364/oe.24.0a1210

1295 Wu, P., Shen, H., Ai, T., & Liu, Y. (2013). Land-surface temperature retrieval at high spatial and
1296 temporal resolutions based on multi-sensor fusion. *International Journal of Digital Earth,*
1297 *6*(sup1), 113-133. https://doi.org/10.1080/17538947.2013.783131

1298 Wulfmeyer, V., Turner, D. D., Baker, B., Banta, R., Behrendt, A., Bonin, T., et al. (2018). A new
1299 research approach for observing and characterizing land-atmosphere feedback. *Bulletin of the*

American Meteorological Society, 99(8), 1639-1667. Article. https://doi.org/10.1175/bams-d-17-0009.1

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. https://doi.org/10.1175/1520-0469(1984)041<0102:TDABUD>2.0.CO;2

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, 232*, 10-22. https://doi.org/10.1016/j.agrformet.2016.07.019

Xu, K., Metzger, S., & Desai, A. R. (2018). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. *Agricultural and Forest Meteorology, 255*, 81-91. https://doi.org/10.1016/j.agrformet.2017.10.011

Xu, K., Sühring, M., Metzger, S., Durden, D., & Desai, A. R. (2020). Can data mining help eddy covariance see the landscape? A large-eddy simulation study. *Boundary-Layer Meteorology, Online First*. https://doi.org/10.1007/s10546-020-00513-0

Xu, Z., Ma, Y., Liu, S., Shi, W., & Wang, J. (2016). Assessment of the energy balance closure under advective conditions and its impact using remote sensing data. *Journal of Applied Meteorology and Climatology, 56*(1), 127-140. https://doi.org/10.1175/JAMC-D-16-0096.1

Younker, D. (2003). *Value Engineering: Analysis And Methodology*. Boca Raton, USA: CRC Press.

Zhang, L., & Pu, Z. (2010). An Observing System Simulation Experiment (OSSE) to assess the impact of Doppler Wind Lidar (DWL) measurements on the numerical simulation of a tropical cyclone. *Advances in Meteorology, 2010*, 743863. https://doi.org/10.1155/2010/743863

Zhang, Y., Ye, X., Wang, S., He, X., Dong, L., Zhang, N., et al. (2020). Large-eddy simulation of traffic-related air pollution at a very high-resolution in a mega-city: Evaluation against mobile sensors and insights for influencing factors. *Atmospheric Chemistry and Physics Discussions, 2020*, 1-16. https://doi.org/10.5194/acp-2020-1168

1326 Ziehn, T., Law, R. M., Rayner, P. J., & Roff, G. (2016). Designing optimal greenhouse gas
1327 monitoring networks for Australia. *Geosci. Instrum. Method. Data Syst., 5*(1), 1-15.
1328 https://doi.org/10.5194/gi-5-1-2016