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]

5   [1]Battelle, National Ecological Observatory Network, 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,
10   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

Correspondence to: Stefan Metzger (smetzger@battelleecology.org)

15   **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

20   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

25   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

30   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.

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

35   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 (Masutani et al., 2010; Hoffman and Atlas, 2016; Atlas et al., 2015) 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

40   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

45   Energy-balance Study Enabled by a High-density Extensive Array of Detectors 2019" (CHEESEHEAD19; Butterworth et al., 2021). 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

50   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

55   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 (Fig. 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

60   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.

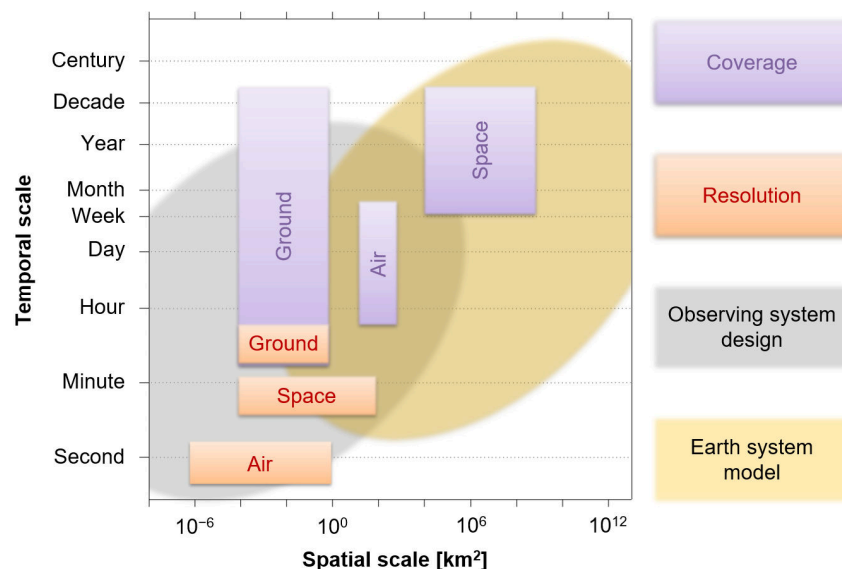



Figure 1. Space-time scope diagram for a surface-atmosphere synthesis observing system at the interface of meso- and microscale meteorology. The observing system consists of a hierarchy of ground-based (Ground), airborne (Air) and space-borne (Space) measurements, shown in relation to two principal approaches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

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 and 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., Eddy, 1974; Gehrke et al., 2019; Cortina and Calaf, 2017). 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.

Previous studies employed data-intensive scientific discovery for post-field OSD assessments (e.g., Kumar et al., 2016; Chu
et al., 2017; Mahecha et al., 2017; Koffi et al., 2013; Montanari et al., 2012; Loescher et al., 2014; 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; 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 et al., 2013a; Metzger,
2018; 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 and Hoffman, 2004; Kaminski et al.,
2012; Keller et al., 2008; Park and Kim, 2020; Ziehn et al., 2016; Lucas et al., 2015; Sulkava et al., 2011; Zhang and Pu,
2010; Lauvaux et al., 2012). 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 (Fig. 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; 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; Sellers et al.,
1988), the Boreal Ecosystem-Atmosphere Study (BOREAS; 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; Beyrich et al., 2002) and LITFASS-2003 (Beyrich and
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 (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 et al., 2013a). 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 and 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

120 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

125 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

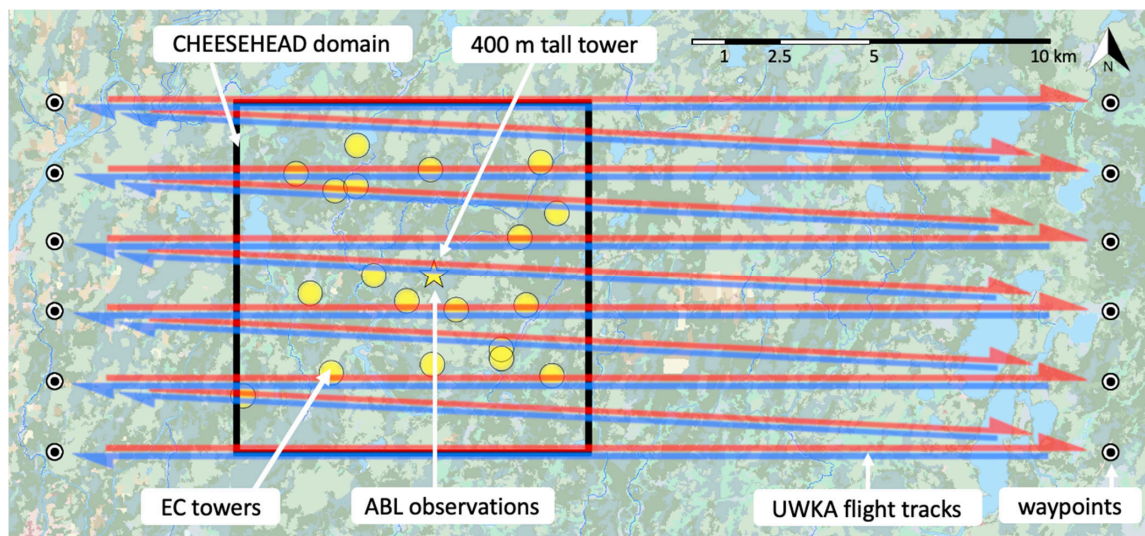
130 The CHEESEHEAD19 study (Butterworth et al., 2021) was to be comprised of a four-month field measurement campaign to investigate how land surface heterogeneity influences energy balance closure. The energy balance closure problem refers to the situation, common in EC measurements, whereby downward energy from radiation and ground heat flux exceeds the measured upward energy from sensible and latent heat fluxes (Foken et al., 2011; Mauder et al., 2020). Previous studies have indicated that heterogeneity does impact energy balance (Stoy et al., 2013; Xu et al., 2016). The CHEESEHEAD19 project

135 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; Mauder et al., 2007b; Gao et al., 2016).

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

140 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; 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 (Xu et al., 2017). The

145 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).

150 **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).**

To measure fluxes (momentum, sensible heat, latent heat, CO_2) across the domain, 20 above-canopy EC towers were proposed

155 to be deployed over a range of vegetation types (Fig. 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; Desai, 1996 -) that has been measuring fluxes at 30 m, 122 m, and 396 m above ground for the past 26 years (1995 - 2020; Desai et al., 2015).

160 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

165 imagery, ECOSTRESS, Landsat, VIIRS and/or GOES (e.g., 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

170 water vapor over the domain (Wu et al., 2016). During each research flight the UWKA would fly along 11 flight tracks (Fig. 2).

For a given flight track the UWKA would first fly outbound at 400 m above ground (Fig. 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 (Fig. 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. (2021).

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 and Schmidt, 1998; Kanda et al., 2004; Sühring et al., 2018; Xu et al., 2020; Margairaz 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., 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 (Fig. 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 (Fig. 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

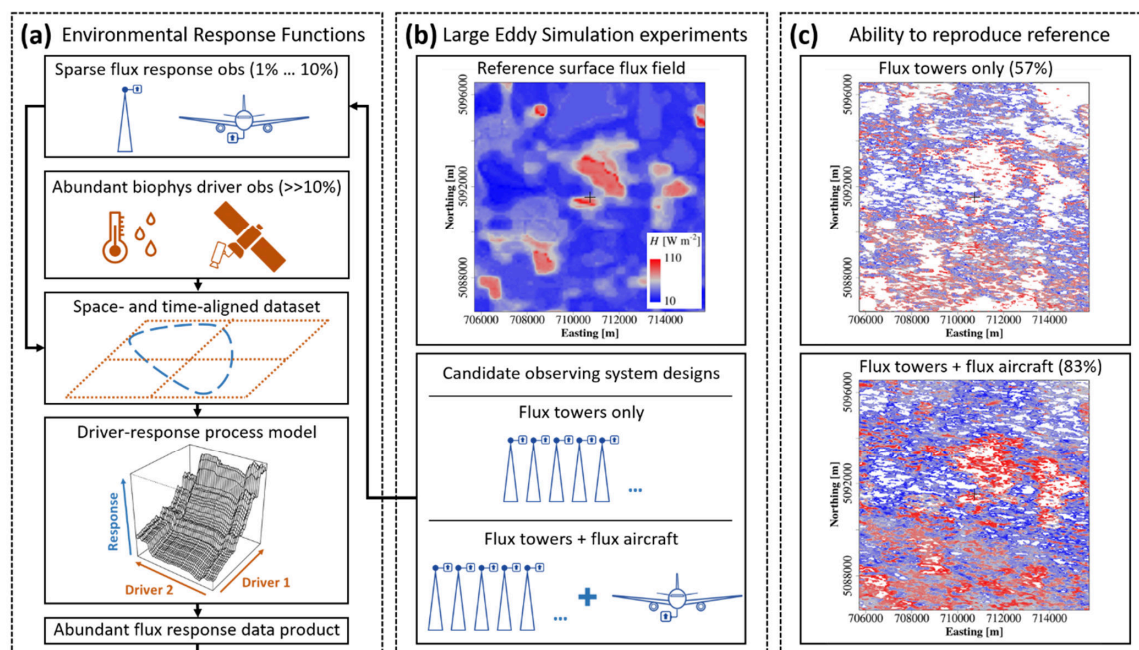



Figure 3. (a) Environmental Response Functions relate tower and airborne eddy-covariance flux 'responses' at very high space-time resolution to meteorological and surface drivers from ground-based, airborne and space-borne measurements. We then use the
205 **extracted relationships to reconstruct fluxes explicitly across the study domain. (b) To simulate different candidate OSDs ahead of the actual field measurement campaign, we used LES to obtain observations in virtual atmospheres over the CHEESEHEAD19 domain. (c): We then benchmarked the different OSD candidates against their ability to reproduce the LES reference in the form of flux grids that ERF reconstructed from the virtual observations alone. Sect. 2.2 provides additional detail. Modified after Butterworth et al. (2021), copyright (2020), with permission from the American Meteorological Society to reuse under the CC BY**
210 **4.0 license (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cc-by-4.0).**

scales of ground-based and space-borne observations (Fig. 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)
215 generating virtual measurements, e.g. in LES, and (iii) benchmarking candidate OSDs (Fig. 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; Fig. 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
220 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 (Fig. 3b). We then super-sampled these

225   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 (Fig. 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–

230   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;

$$EBR = \frac{\langle F_{H,ERF}\rangle + \langle F_{LE,ERF}\rangle}{\langle F_{H,LES}\rangle + \langle F_{LE,LES}\rangle}; \tag{1}$$

        with angle brackets indicating the horizontal average over all grid cells in the study domain, and F_H and F_{LE} the

235   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).

240   ## 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

245   (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

250   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

255 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 (Fig. 2 blue arrows). To create the dataset for assessing H1 we formed the virtual flight patterns by letting multiple aircraft fly simultaneous grids

260 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

265 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

270 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.

275 **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 Deardorff (1980) for the sub-grid model, a fifth-order scheme (Wicker and Skamarock, 2002) to discretize the advection terms, and a third-order Runge–Kutta scheme by Williamson

280 (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

285   grids that Metzger et al. (2013b) 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

290   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

295   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 $10 \times 10\ km$ CHEESEHEAD19 domain. The grid spacing was 25 m in the horizontal directions and 10 m in the vertical direction. The

300   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 $82\ m\ s^{-1}$.

305   **2.5**   **Environmental Response Functions**

ERF incorporates spectral averaging, source area modeling, and machine-learning-based space-time explicit ensembling of environmental observations (Metzger et al., 2013a). 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; Xu et al., 2017). These flux maps comply with observational assumptions that are not typically met from EC measurements alone, such as incorporation

310   of mesoscale flows and spatial representativeness for the $10 \times 10\ km$ CHEESEHEAD19 target domain (Xu et al., 2020; Xu et al., 2018; Metzger, 2018).

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)

315   resolution of the resulting fluxes (Mauder et al., 2007a). 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

320 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 (Metzger et al., 2012; Kljun et al., 2004) 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

325 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

330 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

335 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

340 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

345 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

350 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
355   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.

360

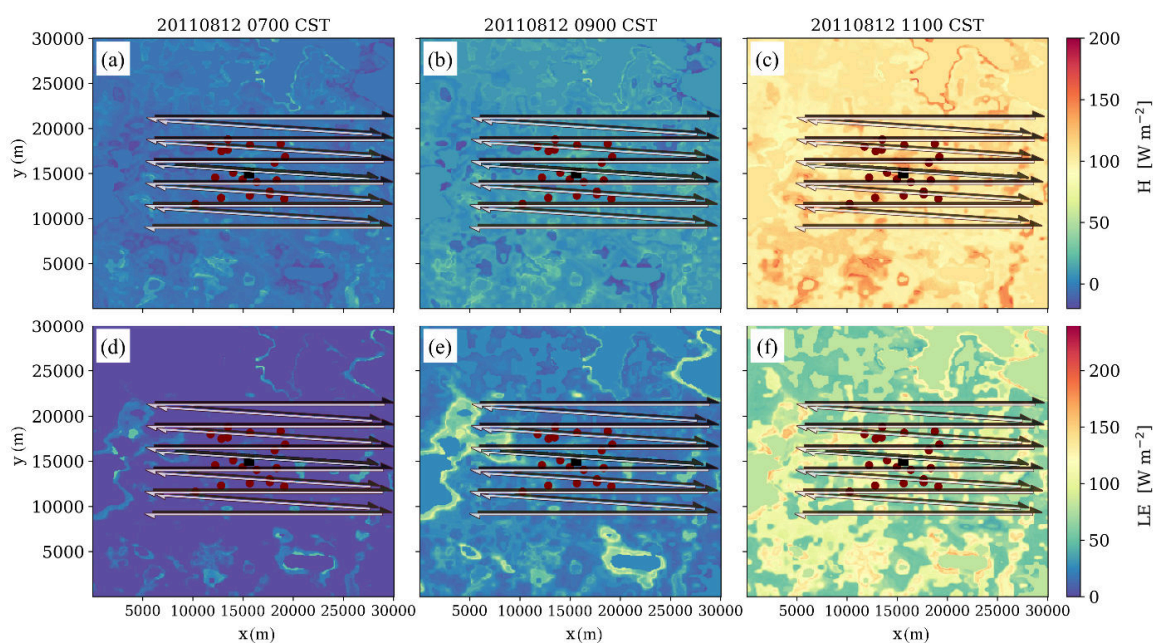



Figure 4. Time-sequence of (a) – (c) the spatially heterogeneous surface sensible heat flux and (d) – (f) latent heat flux prescribed as LES lower boundary conditions. Superimposed red dots indicate candidate EC tower locations, alongside UWKA candidate flight tracks (west-east pattern).

365

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

370   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
375   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.

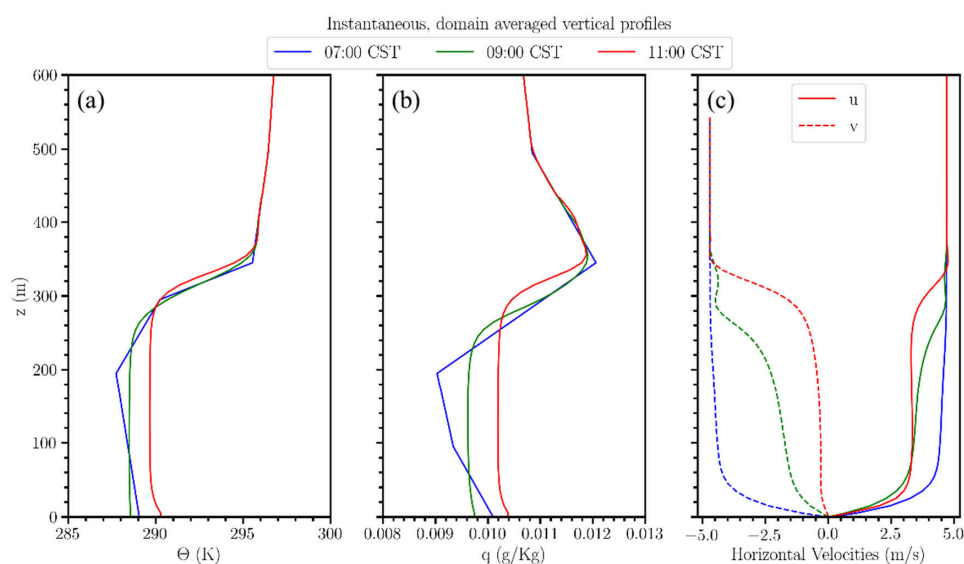



Figure 5. LES domain-averaged vertical profiles of (a) potential temperature, (b) water vapor mixing ratio and (c) horizontal wind
380   **velocities at different simulation times.**

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.
385   Even though these organized structures are not stationary in space due to the slightly changing wind direction (see Fig. 5),
they still can be observed in the temporal average.

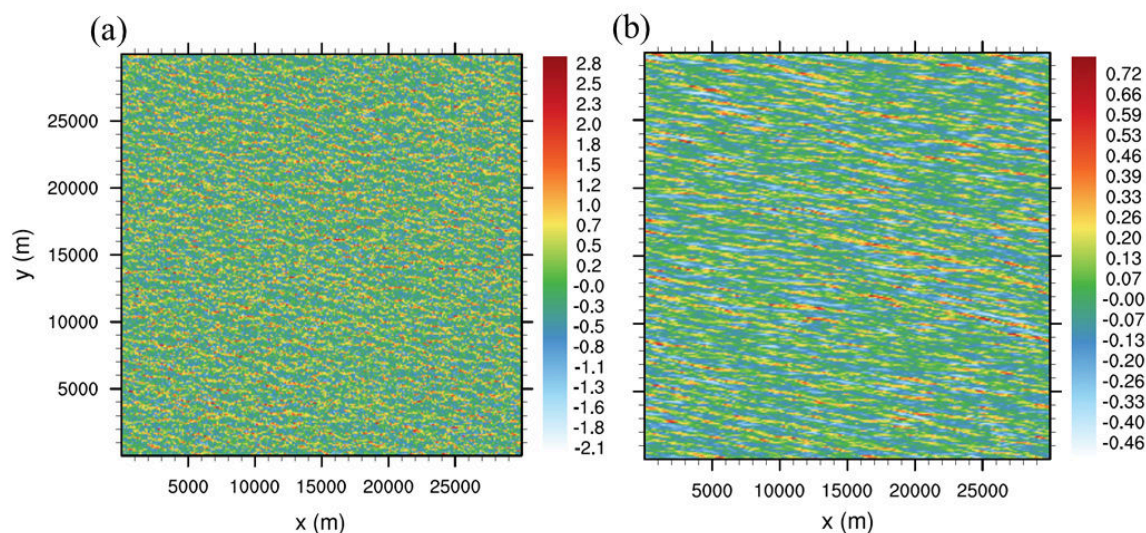



Figure 6. LES x-y cross section of (a) instantaneous and (b) 30-minute time-averaged vertical velocity at 100 m height at 11:00 CST simulation time.

390

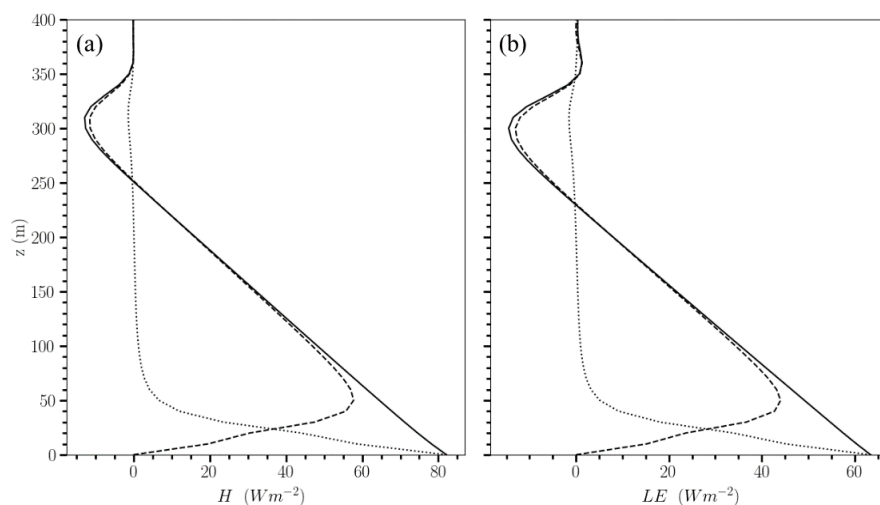



Figure 7. Domain averaged vertical flux profiles of (a) sensible heat and (b) latent heat at 11:00 CST simulation time. The solid lines show the total simulated fluxes, consisting of resolved fluxes (dashed lines) and sub-grid parameterized fluxes (dotted lines).

Figure 7 shows vertical profiles of the domain-averaged sensible and latent heat flux. Both flux profiles display a similar

395 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 Fig. 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.

400 **3.2   ERF retrievals**

To create a space- and time-aligned data set (Fig. 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 (Fig. 8).

405

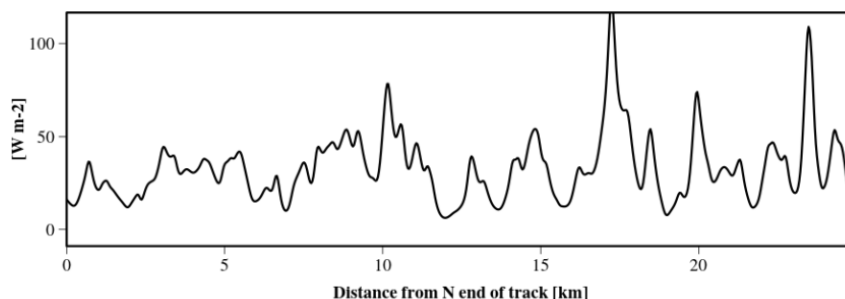



Figure 8. Space-resolved sensible heat flux from high-rate space-scale decomposition of virtual airborne measurements.

Next, ERF relates the time- and space-resolved EC flux responses in the atmosphere to biophysical drivers on the surface

(Fig. 9), such as land surface temperature and land surface moisture. In the present application land surface temperature and

410 land surface moisture are taken from cross-sections at the vertical LES level closest to the surface. In real-world ERF

applications, these are substituted with space-borne and airborne remote sensing data products, or reanalysis data (e.g.,

Serafimovich et al., 2018). This facilitates mathematically rigorous data overlays among response and driver observations at

minute- and decameter-scale. The result is a space-time aligned data set for each virtual EC tower and for each virtual EC

flight track. Both, the tower and airborne EC data sets comprise the same variables in identical units, and were processed to

415 ensure cross-platform compatibility and avoid biases (Sect. 2.5). This allows combining the virtual tower EC results with

corresponding virtual airborne EC results into a single space-time aligned data set for each of the 12 alternative OSDs.

The ERF machine learning component then extracts a driver-response process model from the resulting space- and time-

aligned dataset. Here, we trained a total of 78 machine learning models. These consisted of 13 candidate OSDs × 2 fluxes

(sensible and latent heat) × 3 replicates (to quantify stochastic uncertainty in the response data products). Figure 10 shows

420

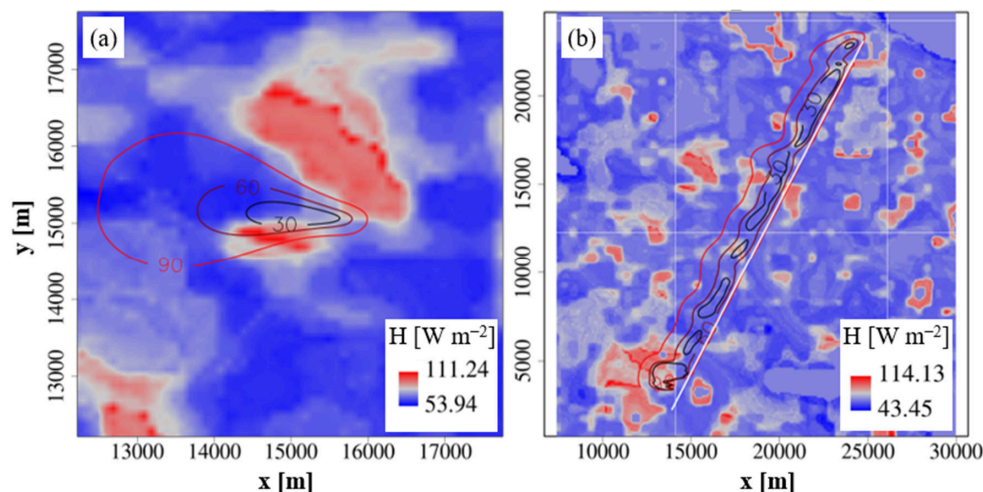



Figure 9. Source area modeling (30%, 60%, 90% contour lines) relates observations across platforms and representations by determining the surface sources of the time- and space-resolved EC fluxes. Here superimposed over the LES sensible heat flux surface forcing (a) Virtual AmeriFlux supersite tower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

425

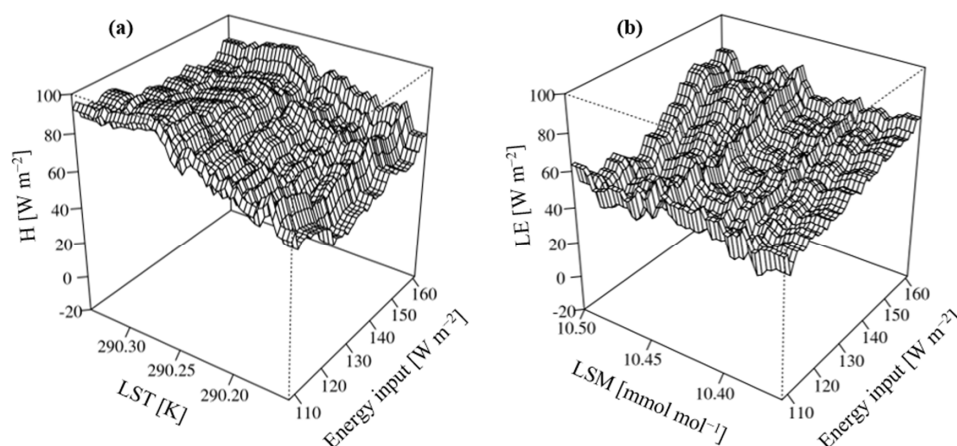



Figure 10. Example ERF response surfaces. (a) Sensible heat flux as a function of energy input and land surface temperature (LST). (b) Latent heat flux as a function of energy input and land surface moisture (LSM). For this visualization, all other drivers are kept at their median value.

430 example driver-response surfaces for sensible and latent heat flux as a function of their principal drivers, energy input, land surface temperature and land surface moisture, respectively. This exemplifies how the turbulent diffusion coefficient connects the drivers to the flux response within the physics-guided ERF model structure. In Fig. 10a the sensible heat flux increases primarily with land surface temperature and secondarily with energy input. The relationship reaches a plateau around 290.3 K which deviates from a one-dimensional, monotonic and linear gradient-flux relationship, indicative of

435 additional feedback processes. Conversely, in Fig. 10b the latent heat flux increases primarily with energy input and secondarily with land surface moisture, with monotonic and approximately linear relationships across the range of drivers. Ultimately, the physics-guided ERF 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-resolution sensible heat flux maps, projected explicitly in space and time across the study domain to the median relative measurement

440 height of the 49 m towers (0.16 of the ABL height; Fig. 11). We projected the flux maps for the tower-only space-time aligned data set (baseline OSD; Fig. 11a), and then separately for each of the 12 joint tower and aircraft space-time aligned data sets (alternative OSDs; Fig. 11c). Now the flux maps that ERF reconstructed from the virtual measurements alone can be compared to the original LES surface flux forcings (Fig. 11b).

445

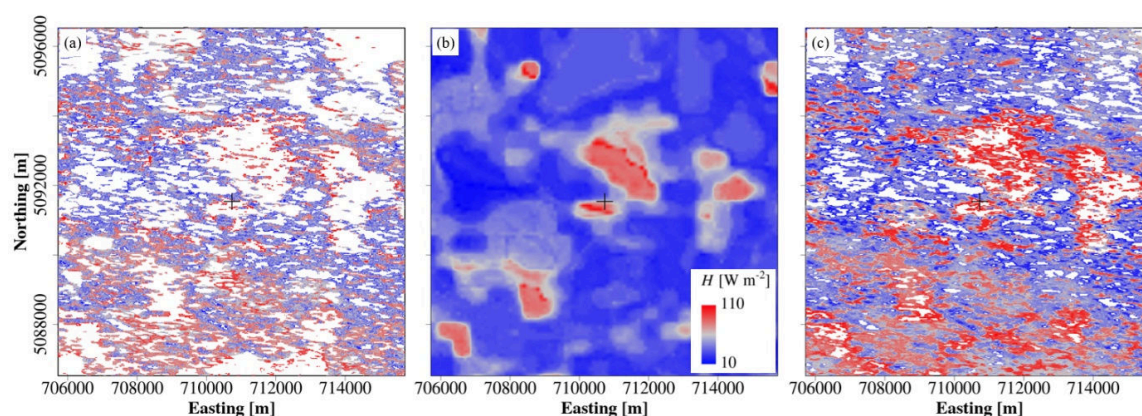



Figure 11. Example ERF response data products: sensible heat flux maps independently derived for (a) the tower-only space-time aligned data set and (c) for the joint tower and aircraft space-time aligned data set, alongside (b) the LES reference surface flux field.

450 **3.3   Evaluation of design hypotheses**

The ERF-derived flux maps alongside the LES surface forcing in Fig. 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

455   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

460   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).

465   **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%,

470   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, Fig. 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, Fig. 12b shows less

overlapping source areas along the flight track of wind-perpendicular flights, with the comparatively weaker cross-wind

475   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.

480   Furthermore, at the virtual aircraft flight height of 100 m the time-averaged vertical wind cross-section in Fig. 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

(Mauder et al., 2020). These structures consist of more spatially-expansive though weaker subsidence zones and more

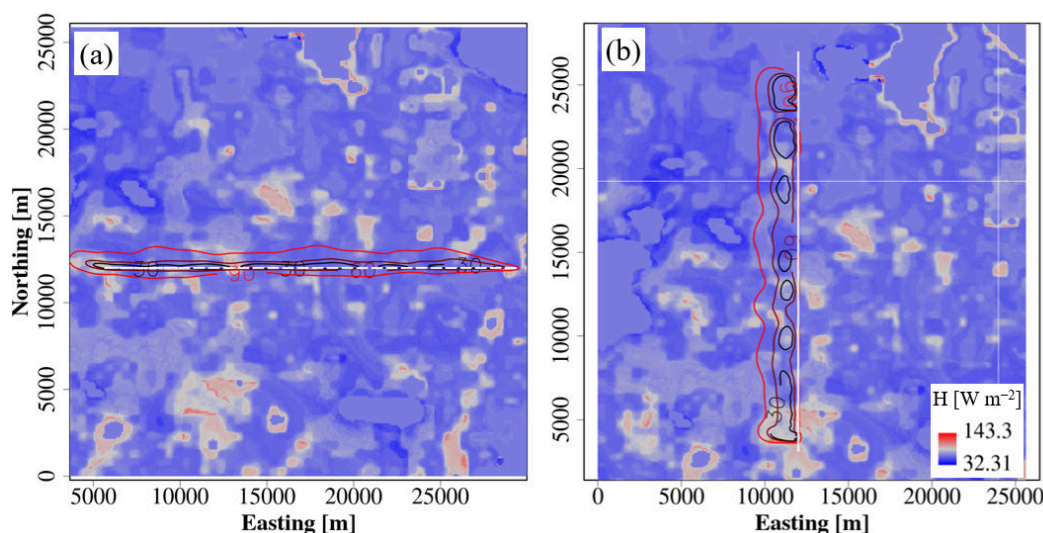



Figure 12. Example virtual flight tracks and their 30%, 60% and 90% source area contours superimposed over the LES sensible heat flux surface forcing (W m⁻²). (a) Wind-parallel flights sample source areas that are elongated along the flight track leading to considerable overlap. (b) Wind-perpendicular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

spatially-limited though stronger convection zones (Kanda et al., 2004; Lenschow and Stankov, 1986; Petty, 2020; Etling and Brown, 1993; Moeng and Rotunno, 1990). 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.

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, 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.

510 **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

515 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.

520

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

525

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 (Fig. 13). Flying the numbered waypoint in

530   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

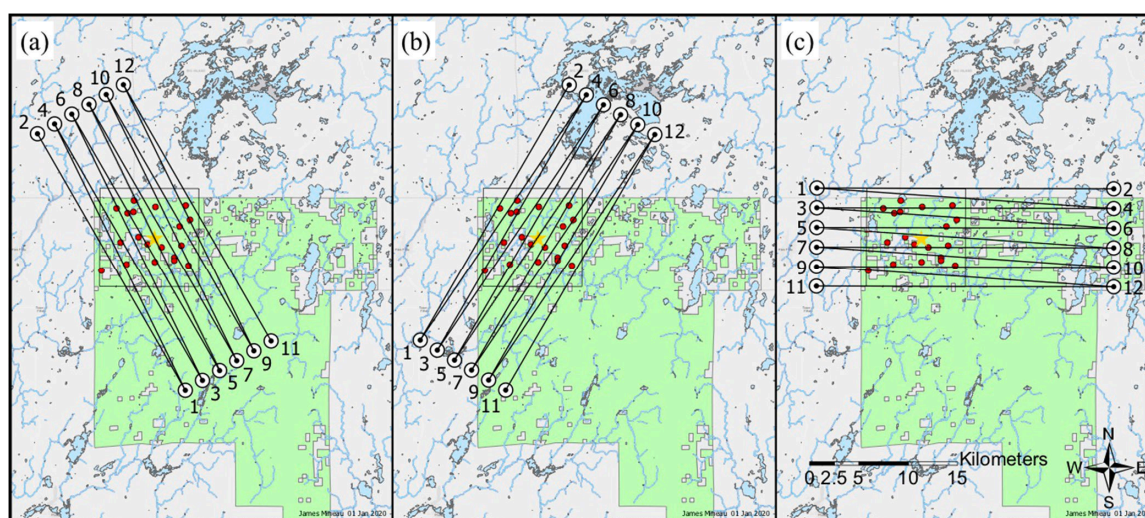
three additional alternating flight sequences SE2, SW2 and WE2 with 150°, 210°, and 270° azimuth of the parallel tracks,

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

535   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 (Fig. 13c).

540   **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.**

545   To support daily flight planning we distilled the six alternating flight sequences into a flight planning wind rose (Fig. 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

550   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.

555

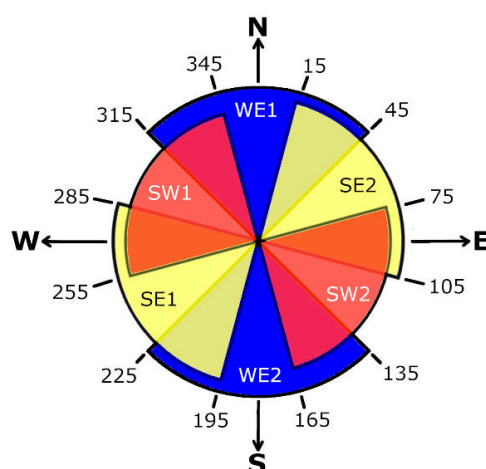



Figure 14. Flight planning wind rose to choose the appropriate flight sequence as a function of the forecasted mean wind direction. Owing to rotational symmetry, the wind sector for each flight sequence overlaps with each of its two neighbors by 30°. This provides a margin for accommodating changing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

4 Discussion

560 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

565 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 objectives numerically. Furthermore, we identified two specific design hypotheses that we postulate the science objectives, and hence optimality criteria to be sensitive to (Sect. 2.3; H1

570 track angle on mean wind, H2 fine vs. coarse flight sequence). CHEESEHEAD19's first science objective O1 is to show that higher surface heterogeneity promotes energy transport in atmospheric mesoscale eddies. Our ability to address this science

objective increases with the truthful reproduction of CR1 surface flux spatial coverage and CR3 spatial patterning. LES-ERF allowed us to assess changes in these criteria resulting from the different OSDs by comparing ERF flux map reproductions to the original LES surface flux forcing. We found that CR1 spatial coverage is largely insensitive to either H1 track angle on

575 mean wind (Table 1), and H2 fine vs. coarse flight sequence (Table 2). Conversely, CR3 spatial patterning proved to be highly sensitive to H1 track angle on mean wind. Track angles in the range of 90°±45° on the mean wind yielded double to triple improvements over wind-parallel flights (Table 1). Similarly, we showed that CR3 spatial patterning is sensitive to H2 fine vs. coarse flight sequence (Table 2). The finely spaced "Alternating" sequence yielded the highest and most consistent spatial patterning improvements of about 50% over the other flight sequences.

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

585 insensitive to H2 fine vs. coarse flight sequence (Table 2).

Finally, CHEESEHEAD19's third science objective O3 aims to demonstrate that ERF yields representative fluxes at model grid scale regardless of mesoscale eddies. ERF's ability to reproduce the surface flux is thus directly related to the combination of all criteria discussed above, CR1 spatial coverage, CR2 energy balance ratio, and CR3 spatial patterning. From aggregating over all optimality criteria into a single score, we demonstrated that overall improvement is highly

590 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

595 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

600 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

605 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

610   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

615   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

620   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

625   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

630   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; Xu et al., 2020; Sühring et al., 2018).

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

635   planning covered 480 flight legs during 72 hours of flight time and three CHEESEHEAD19 intensive observation periods (Butterworth et al., 2021). 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,
640 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 (Sühring and Raasch, 2013; Kanda et al., 2004; Deardorff, 1972; Wyngaard and Brost, 1984; Etling and
645 Brown, 1993). Indeed, LES have become useful to contextualize observational phenomena with increasingly complex
feedbacks, including natural resources such as air quality (Khan et al., 2020; 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. 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
650 LES-ERF symbiosis for designing real-world field measurements. Using LES for OSD has been a rare application to date
(Gehrke et al., 2019; Cortina and Calaf, 2017), 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 (Hoffman and Atlas, 2016; Masutani et al., 2010; Lucas et al., 2015; Park and Kim, 2020; Hargrove and
655 Hoffman, 2004; Zhang and 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.

660 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 (Wulfmeyer et al., 2018; Helbig et al., 2020) and soil
and biometric observations (Metzger et al., 2019b). 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
665 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 et al., 2019a). 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
670 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

675   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%

680   and ~60%, respectively. The main labor drivers were study conceptualization and setup including data acquisition for LES boundary conditions. It is possible to perform these steps well in advance, e.g. to reduce the LES-ERF effort between grant receipt and field measurements, which is typically also a period with high-demand for overall coordination. The 50 h spent on ERF interpretation also included active dialogue and iteration with the flight crew, resulting in balanced resources for airborne operation.

685

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

Table 4 also shows how LES and ERF differed in their computational needs. LES demanded high-performance computing

690   with 230,000 CPU hours and up to 120 TB memory, which we primarily performed on the US National Center for

Atmospheric Research Cheyenne supercomputer (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cheyenne). Conversely, ERF required a high-throughput computing architecture, for which we primarily used the US National Science Foundation's CyVerse open science workspace (https://w3id.org/smetzger/Metzger-et-al_2021_OSSE-LES-ERF/cyverse). Overall, the strong data requirements of ERF, including use of high-frequency EC data, currently drive

695 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 (Sogachev et al., 2002; Santiago et al., 2010). 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

700 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

705 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 (Fig. 1). However, data

710 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 (Wu et al., 2013; Pincebourde and Salle, 2020).

715 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 and Foken, 2014). However, e.g.Bertoldi et al. (2013); 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.

720 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

725   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

730   ("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

735   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

740   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.

Code availability. 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

745   developed into public code repositories.

Data availability. 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

750   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.

Author contribution. Conception and conceptualization: SM. Methodology: SM, DD, SP, MS, DP, KX, AD. Software: SM, DD, SP, MS. Validation: SM, DD, SP, MS. Formal analysis: SM, DD, SP, MS, KX. Investigation: SM, DD, SP, MS, AD.

755 Resources: SM, DD, SP, MS, AD. Data curation: SM, DD, AD. Original draft: SM, DD, SP, MS, BB, MM, LW. Review and editing: SM, DD, SP, MS, BB, CF, MM, DP, LW, AD. Visualization: SM, DD, SP, BB. Supervision: SM, AD. Project administration: SM, CF, AD. Funding acquisition: SM, AD.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The National Ecological Observatory Network is a program sponsored by the National Science

760 Foundation and operated under cooperative agreement by Battelle Memorial Institute. This material is based in part upon work supported by the National Science Foundation through the NEON Program (grant DBI-0752017) and the CHEESEHEAD19 project (grant AGS-1822420). 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. LW's and MM's contributions were supported by Deutsche Forschungsgemeinschaft (DFG; grant #406980118).

765 References

Atlas, R., Hoffman, R. N., Ma, Z., Emmitt, G. D., Wood, S. A., Jr., Greco, S., Tucker, S., Bucci, L., Annane, B., Hardesty, R. M., and Murillo, S.: Observing System Simulation Experiments (OSSEs) to evaluate the potential impact of an Optical Autocovariance Wind Lidar (OAWL) on numerical weather prediction, J. Atmos. Oceanic Technol., 32, 1593-1613, doi:10.1175/jtech-d-15-0038.1, 2015.

770 Avissar, R., and Schmidt, T.: 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, 2666-2689, 10.1175/1520-0469(1998)055<2666:aeotsa>2.0.co;2, 1998.

Bertoldi, G., Kustas, W., and Albertson, J.: Evaluating source area contributions from aircraft flux measurements over heterogeneous land using large-eddy simulation, Boundary Layer Meteorol., 147, 261-279, doi:10.1007/s10546-012-9781-y,

775 2013.

Beyrich, F., Herzog, H. J., and Neisser, J.: The LITFASS project of DWD and the LITFASS-98 experiment: The project strategy and the experimental setup, Theor. Appl. Climatol., 73, 3-18, doi:10.1007/s00704-002-0690-8, 2002.

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

780   Brown, J. R., and Fehige, Y.: Thought experiments, in: The Stanford Encyclopedia of Philosophy, Winter 2019 ed., edited by: Zalta, E. N., Metaphysics Research Lab, Stanford University, Stanford, USA, https://plato.stanford.edu/archives/win2019/entries/thought-experiment/, 2019.

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

785   Butterworth, B. J., Desai, A. R., Metzger, S., Townsend, P. A., Schwartz, M. D., Petty, G. W., Mauder, M., Vogelmann, H., Andresen, C. G., Augustine, T. J., Bertram, T. H., Brown, W. O. J., Buban, M., Cleary, P., Durden, D. J., Florian, C. R., Iglinski, T. J., Kruger, E. L., Lantz, K., Lee, T. R., Meyers, T. P., Mineau, J. K., Olson, E. R., Oncley, S. P., Paleri, S., Pertzborn, R. A., Pettersen, C., Plummer, D. M., Riihimaki, L., Guzman, E. R., Sedlar, J., Smith, E. N., Speidel, J., Stoy, P. C., Sühring, M., Thom, J. E., Turner, D. D., Vermeuel, M. P., Wagner, T. J., Wang, Z., Wanner, L., White, L. D., Wilczak, J.

790   M., Wright, D. B., and Zheng, T.: Connecting land-atmosphere interactions to surface heterogeneity in CHEESEHEAD19, Bull. Am. Meteorol. Soc., 102, E421 - E445, doi:10.1175/bams-d-19-0346.1, 2021.

Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-P., and Foken, T.: Extension of the averaging time in eddy-covariance measurements and its effect on the energy balance closure, Boundary Layer Meteorol., 152, 303-327, doi:10.1007/s10546-014-9922-6, 2014.

795   Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., Black, T. A., Arain, M. A., Bourque, C. P. A., Flanagan, L. B., Lafleur, P. M., McCaughey, J. H., and Wofsy, S. C.: Assessing eddy-covariance flux tower location bias across the Fluxnet-Canada Research Network based on remote sensing and footprint modelling, Agric. For. Meteorol., 151, 87-100, doi:10.1016/j.agrformet.2010.09.005, 2011.

Chu, H., Baldocchi, D. D., John, R., Wolf, S., and Reichstein, M.: Fluxes all of the time? A primer on the temporal

800   representativeness of FLUXNET, Journal of Geophysical Research: Biogeosciences, 122, 289-307, doi:10.1002/2016jg003576, 2017.

Cortina, G., and Calaf, M.: Turbulence upstream of wind turbines: A large-eddy simulation approach to investigate the use of wind lidars, Renewable Energy, 105, 354-365, doi:10.1016/j.renene.2016.12.069, 2017.

Deardorff, J. W.: Numerical investigation of neutral and unstable planetary boundary layers, Journal of the Atmospheric

805   Sciences, 29, 91-115, doi:10.1175/1520-0469(1972)029<0091:nionau>2.0.co;2, 1972.

Deardorff, J. W.: Stratocumulus-capped mixed layers derived from a three-dimensional model, Boundary Layer Meteorol., 18, 495-527, doi:10.1007/bf00119502, 1980.

Desai, A.: AmeriFlux US-PFa Park Falls/WLEF, Dataset, doi:10.17190/AMF/1246090, 1996 -

Desai, A. R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., Andrews, A. E., Cook, B. D., King, J. Y., and

810   Kolka, R.: Landscape-level terrestrial methane flux observed from a very tall tower, Agric. For. Meteorol., 201, 61-75, doi:10.1016/j.agrformet.2014.10.017, 2015.

Desjardins, R. L., Worth, D. E., Pattey, E., VanderZaag, A., Srinivasan, R., Mauder, M., Worthy, D., Sweeney, C., and Metzger, S.: The challenge of reconciling bottom-up agricultural methane emissions inventories with top-down measurements, Agric. For. Meteorol., 248, 48-59, doi:10.1016/j.agrformet.2017.09.003, 2018.

815 Eddy, A.: An approach to the design of meteorological field experiments, Monthly Weather Review, 102, 702-707, doi:10.1175/1520-0493(1974)102<0702:aattdo>2.0.co;2, 1974.

Etling, D., and Brown, R. A.: Roll vortices in the planetary boundary layer: A review, Boundary Layer Meteorol., 65, 215-248, doi:10.1007/bf00705527, 1993.

Foken, T., Aubinet, M., Finnigan, J. J., Leclerc, M. Y., Mauder, M., and Paw U, K. T.: Results of a panel discussion about

820 the energy balance closure correction for trace gases, Bull. Am. Meteorol. Soc., 92, ES13-ES18, doi:10.1175/2011bams3130.1, 2011.

Foken, T.: Micrometeorology, 2 ed., Springer, Berlin, Heidelberg, 362 pp., 2017.

Gao, Z., Liu, H., Russell, E. S., Huang, J., Foken, T., and Oncley, S. P.: Large eddies modulating flux convergence and divergence in a disturbed unstable atmospheric surface layer, Journal of Geophysical Research: Atmospheres, 121, 1475-

825 1492, doi:10.1002/2015jd024529, 2016.

Gehrke, K., Böske, L., and Beyrich, F.: Analyse verschiedener Doppler-Lidar Scan-Strategien zur Ableitung des mittleren Windvektors in der konvektiven Grenzschicht auf Basis von Large-Eddy-Simulationen, DACH conference of German, Austrian and Swiss meteorologists, Garmisch-Partenkirchen, Germany, 18 - 22 March, 2019.

Halldin, S., Gryning, S. E., Gottschalk, L., Jochum, A., Lundin, L. C., and Van de Griend, A. A.: Energy, water and carbon

830 exchange in a boreal forest landscape - NOPEX experiences, Agric. For. Meteorol., 98-99, 5-29, doi:10.1016/S0168-1923(99)00148-3, 1999.

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

Helbig, M., Gerken, T., Beamesderfer, E., Baldocchi, D. D., Banerjee, T., Biraud, S. C., Brunsell, N. A., Burns, S. P.,

835 Butterworth, B., Chan, W. S., Desai, A. R., Fuentes, J. D., Hollinger, D. Y., Kljun, N., Mauder, M., Novick, K. A., Perkins, J. M., Rey-Sanchez, C., Scott, R. L., Seyednasrollah, B., Stoy, P. C., Sullivan, R. C., Vilà-Guerau de Arellano, J., Wharton, S., Yi, C., and Richardson, A. D.: White Paper: Understanding land-atmosphere interactions through tower-based flux and continuous atmospheric boundary layer measurements, AmeriFlux Management Project, Berkeley, CA, U.S.A., 47 pp., 2020.

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

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

Hoffman, R. N., and Atlas, R.: Future observing system simulation experiments, Bull. Am. Meteorol. Soc., 97, 1601-1616,

845 doi:10.1175/bams-d-15-00200.1, 2016.

Kaharabata, S. K., Schuepp, P. H., Ogunjemiyo, S., Shen, S., Leclerc, M. Y., Desjardins, R. L., and MacPherson, J. I.: Footprint considerations in BOREAS, J. Geophys. Res. Atmos., 102, 29113-29124, doi:10.1029/97JD02559, 1997.

Kaminski, T., Rayner, P. J., Voßbeck, M., Scholze, M., and Koffi, E.: Observing the continental-scale carbon balance: Assessment of sampling complementarity and redundancy in a terrestrial assimilation system by means of quantitative
850   network design, Atmos. Chem. Phys., 12, 7867-7879, doi:10.5194/acp-12-7867-2012, 2012.

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

Keller, M., Schimel, D. S., Hargrove, W. W., and Hoffman, F. M.: A continental strategy for the National Ecological Observatory Network, Frontiers in Ecology and the Environment, 6, 282-284, doi:10.1890/1540-
855   9295(2008)6[282:acsftn]2.0.co;2, 2008.

Khan, B., Banzhaf, S., Chan, E. C., Forkel, R., Kanani-Sühring, F., Ketelsen, K., Kurppa, M., Maronga, B., Mauder, M., Raasch, S., Russo, E., Schaap, M., and Sühring, M.: Development of an atmospheric chemistry model coupled to the PALM model system 6.0: Implementation and first applications, Geosci. Model Dev. Discuss., 2020, 1-34, doi:10.5194/gmd-2020-286, 2020.

860   Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple parameterisation for flux footprint predictions, Boundary Layer Meteorol., 112, 503-523, doi:10.1023/B:BOUN.0000030653.71031.96, 2004.

Koffi, E. N., Rayner, P. J., Scholze, M., Chevallier, F., and Kaminski, T.: Quantifying the constraint of biospheric process parameters by CO_2 concentration and flux measurement networks through a carbon cycle data assimilation system, Atmos. Chem. Phys., 13, 10555-10572, doi:10.5194/acp-13-10555-2013, 2013.

865   Kohnert, K., Serafimovich, A., Metzger, S., Hartmann, J., and Sachs, T.: Strong geologic methane emissions from discontinuous terrestrial permafrost in the Mackenzie Delta, Canada, Scientific Reports, 7, 5828, doi:10.1038/s41598-017-05783-2, 2017.

Kumar, J., Hoffman, F. M., Hargrove, W. W., and Collier, N.: Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements, Earth Syst. Sci. Data Discuss., 2016, 1-25, doi:10.5194/essd-
870   2016-36, 2016.

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

Leclerc, M. Y., and Foken, T.: Footprints in micrometeorology and ecology, 1st ed., Springer, Berlin, Heidelberg, Germany,
875   239 pp., 2014.

Lenschow, D. H., and Stankov, B. B.: Length scales in the convective boundary layer, Journal Of The Atmospheric Sciences, 43, 1198-1209, doi:10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2, 1986.

Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics?, J. Atmos. Oceanic Technol., 11, 661-673, doi:10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2, 1994.

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

Lucas, D. D., Yver Kwok, C., Cameron-Smith, P., Graven, H., Bergmann, D., Guilderson, T. P., Weiss, R., and Keeling, R.: Designing optimal greenhouse gas observing networks that consider performance and cost, Geosci. Instrum. Method. Data Syst., 4, 121-137, doi:10.5194/gi-4-121-2015, 2015.

885 Mahecha, M. D., Gans, F., Sippel, S., Donges, J. F., Kaminski, T., Metzger, S., Migliavacca, M., Papale, D., Rammig, A., and Zscheischler, J.: Detecting impacts of extreme events with ecological in situ monitoring networks, Biogeosciences, 14, 4255-4277, doi:10.5194/bg-14-4255-2017, 2017.

Mahrt, L.: The bulk aerodynamic formulation over heterogeneous surfaces, Boundary Layer Meteorol., 78, 87-119, doi:10.1007/BF00122488, 1996.

890 Margairaz, F., Pardyjak, E. R., and Calaf, M.: Surface thermal heterogeneities and the atmospheric boundary layer: The relevance of dispersive fluxes, Boundary Layer Meteorol., 175, 369-395, doi:10.1007/s10546-020-00509-w, 2020.

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., Ketelsen, K., Letzel, M. O., Sühring, M., and Raasch, S.: 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, 2515-2551, doi:10.5194/gmd-8-

895 2515-2015, 2015.

Maronga, B., Banzhaf, S., Burmeister, C., Esch, T., Forkel, R., Fröhlich, D., Fuka, V., Gehrke, K. F., Geletič, J., Giersch, S., Gronemeier, T., Groß, G., Heldens, W., Hellsten, A., Hoffmann, F., Inagaki, A., Kadasch, E., Kanani-Sühring, F., Ketelsen, K., Khan, B. A., Knigge, C., Knoop, H., Krč, P., Kurppa, M., Maamari, H., Matzarakis, A., Mauder, M., Pallasch, M., Pavlik, D., Pfafferott, J., Resler, J., Rissmann, S., Russo, E., Salim, M., Schrempf, M., Schwenkel, J., Seckmeyer, G.,

900 Schubert, S., Sühring, M., von Tils, R., Vollmer, L., Ward, S., Witha, B., Wurps, H., Zeidler, J., and Raasch, S.: Overview of the PALM model system 6.0, Geosci. Model Dev., 13, 1335-1372, doi:10.5194/gmd-13-1335-2020, 2020.

Mason, P. J.: The formation of areally-averaged roughness lengths, Q. J. R. Meteorolog. Soc., 114, 399-420, doi:10.1002/qj.49711448007, 1988.

Masutani, M., Woollen, J. S., Lord, S. J., Emmitt, G. D., Kleespies, T. J., Wood, S. A., Greco, S., Sun, H., Terry, J., Kapoor,

905 V., Treadon, R., and Campana, K. A.: Observing system simulation experiments at the National Centers for Environmental Prediction, Journal of Geophysical Research: Atmospheres, 115, doi:10.1029/2009JD012528, 2010.

Mauder, M., Desjardins, R. L., and MacPherson, I.: Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem, J. Geophys. Res. Atmos., 112, D13112, doi:10.1029/2006JD008133, 2007a.

Mauder, M., Jegede, O. O., Okogbue, E. C., Wimmer, F., and Foken, T.: Surface energy balance measurements at a tropical

910 site in West Africa during the transition from dry to wet season, Theor. Appl. Climatol., 89, 171-183, doi:10.1007/s00704-006-0252-6, 2007b.

Mauder, M., Foken, T., and Cuxart, J.: Surface-energy-balance closure over land: A review, Boundary Layer Meteorol., doi:10.1007/s10546-020-00529-6, 2020.

Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance
915   flux measurements with a weight-shift microlight aircraft, Atmos. Meas. Tech., 5, 1699-1717, doi:10.5194/amt-5-1699-
2012, 2012.

Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón y Widemann, B., Neidl, F., Schäfer, K., Wieneke,
S., Zheng, X. H., Schmid, H. P., and Foken, T.: Spatially explicit regionalization of airborne flux measurements using
environmental response functions, Biogeosciences, 10, 2193-2217, doi:10.5194/bg-10-2193-2013, 2013a.

920   Metzger, S., Xu, K., Desai, A. R., Taylor, J. R., Kljun, N., Schneider, D., Kampe, T., and Fox, A.: Spatio-temporal
rectification of tower-based eddy-covariance flux measurements for consistently informing process-based models, 46th AGU
annual Fall Meeting, San Francisco, U.S.A., 9 - 13 December, 2013b.

Metzger, S.: Surface-atmosphere exchange in a box: Making the control volume a suitable representation for in-situ
observations, Agric. For. Meteorol., 255, 68-80, doi:10.1016/j.agrformet.2017.08.037, 2018.

925   Metzger, S., Ayres, E., Desai, A., Durden, D., Florian, C., Lee, R., Lunch, C., Luo, H., Pingintha-Durden, N., Roberti, J. A.,
SanClements, M., Sturtevant, C., Xu, K., and Zulueta, R. C.: Synthesized observations and processes for plot- to landscape-
scale research, in: NCAR and NEON Town Hall TH13M, 52nd AGU annual Fall Meeting, San Francisco, U.S.A., 9 - 13
December, 2019a.

Metzger, S., Ayres, E., Durden, D., Florian, C., Lee, R., Lunch, C., Luo, H., Pingintha-Durden, N., Roberti, J. A.,
930   SanClements, M., Sturtevant, C., Xu, K., and Zulueta, R. C.: From NEON field sites to data portal: a community resource for
surface–atmosphere research comes online, Bull. Am. Meteorol. Soc., 100, 2305-2325, doi:10.1175/bams-d-17-0307.1,
2019b.

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

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

Munger, J. W., Loescher, H. W., and Luo, H. Y.: Measurement, tower, and site design considerations, in: Eddy covariance:
A practical guide to measurement and data analysis, edited by: Aubinet, M., Vesala, T., and Papale, D., Springer, Dordrecht,
940   Heidelberg, London, New York, 21-58, 2012.

Norton, J.: Thought experiments in Einstein's work, in: Thought Experiments in Science and Philosophy, 1st ed., edited by:
Horowitz, T., and Massey, G. J., Rowman & Littlefield, Lanham, USA, 335, 1991.

Paleri, S., Desai, A. R., Metzger, S., Mauder, M., Plummer, D. M., and Wanner, L.: Mesoscale flux contributions to surface-
atmosphere interactions across a heterogeneous mid-latitude landscape, 52nd AGU annual Fall Meeting, San Francisoco,
945   U.S.A., 9 - 13 December, 2019.

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

Park, J., and Kim, H. M.: Design and evaluation of CO_2 observation network to optimize surface CO_2 fluxes in Asia using observation system simulation experiments, Atmos. Chem. Phys., 20, 5175-5195, doi:10.5194/acp-20-5175-2020, 2020.

950   Park, R.: Value engineering: A plan for invention, CRC Press, Boca Raton, USA, 352 pp., 1998.

Petty, G. W.: 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, doi:10.5194/amt-2020-235, 2020.

Pincebourde, S., and Salle, A.: On the importance of getting fine-scale temperature records near any surface, Global Change Biol., 26, 6025-6027, doi:10.1111/gcb.15210, 2020.

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

Santiago, J. L., Dejoan, A., Martilli, A., Martin, F., and Pinelli, A.: 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 Meteorol., 135, 109-132,

960   doi:10.1007/s10546-010-9466-3, 2010.

Schmid, H. P.: Experimental design for flux measurements: Matching scales of observations and fluxes, Agric. For. Meteorol., 87, 179-200, doi:10.1016/s0168-1923(97)00011-7, 1997.

Schröter, M., Bange, J., and Raasch, S.: Simulated airborne flux measurements in a LES generated convective boundary layer, Boundary Layer Meteorol., 95, 437-456, doi:10.1023/a:1002649322001, 2000.

965   Schuepp, P. H., Leclerc, M. Y., MacPherson, J. I., and Desjardins, R. L.: Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation, Boundary Layer Meteorol., 50, 355-373, doi:10.1007/BF00120530, 1990.

Sellers, P., Hall, F., Margolis, H., Kelly, B., Baldocchi, D., den Hartog, G., Cihlar, J., Ryan, M. G., Goodison, B., Crill, P., Ranson, K. J., Lettenmaier, D., and Wickland, D. E.: The Boreal Ecosystem-Atmosphere Study (BOREAS): An overview and early results from the 1994 field year, Bull. Am. Meteorol. Soc., 76, 1549-1577, doi:10.1175/1520-

970   0477(1995)076<1549:tbesao>2.0.co;2, 1995.

Sellers, P. J., Hall, F. G., Asrar, G., Strebel, D. E., and Murphy, R. E.: The First ISLSCP Field Experiment (FIFE), Bull. Am. Meteorol. Soc., 69, 22-27, doi:10.1175/1520-0477(1988)069<0022:tfife>2.0.co;2, 1988.

Serafimovich, A., Metzger, S., Hartmann, J., Kohnert, K., Zona, D., and Sachs, T.: Upscaling surface energy fluxes over the North Slope of Alaska using airborne eddy-covariance measurements and environmental response functions, Atmos. Chem.

975   Phys., 18, 10007-10023, doi:10.5194/acp-18-10007-2018, 2018.

Sogachev, A., Menzhulin, G. V., Heimann, M., and Lloyd, J.: A simple three-dimensional canopy – planetary boundary layer simulation model for scalar concentrations and fluxes, Tellus B: Chemical and Physical Meteorology, 54, 784-819, doi:10.3402/tellusb.v54i5.16729, 2002.

Squires, F. A., Nemitz, E., Langford, B., Wild, O., Drysdale, W. S., Acton, W. J. F., Fu, P., Grimmond, C. S. B., Hamilton,

980   J. F., Hewitt, C. N., Hollaway, M., Kotthaus, S., Lee, J., Metzger, S., Pingintha-Durden, N., Shaw, M., Vaughan, A. R.,

Wang, X., Wu, R., Zhang, Q., and Zhang, Y.: Measurements of traffic-dominated pollutant emissions in a Chinese megacity, Atmos. Chem. Phys., 20, 8737-8761, doi:10.5194/acp-20-8737-2020, 2020.

Steinfeld, G., Letzel, M. O., Raasch, S., Kanda, M., and Inagaki, A.: Spatial representativeness of single tower measurements and the imbalance problem with eddy-covariance fluxes: Results of a large-eddy simulation study, Boundary Layer Meteorol., 123, 77-98, doi:10.1007/s10546-006-9133-x, 2007.

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F., and Varlagin, A.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity, Agric. For. Meteorol., 171–172, 137-152, doi:10.1016/j.agrformet.2012.11.004, 2013.

Sühring, M., and Raasch, S.: 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 Meteorol., 1-23, doi:10.1007/s10546-013-9822-1, 2013.

Sühring, M., Metzger, S., Xu, K., Durden, D., and Desai, A.: Trade-offs in flux disaggregation: a large-eddy simulation study, Boundary Layer Meteorol., 170, 69-93, doi:10.1007/s10546-018-0387-x, 2018.

Sulkava, M., Luyssaert, S., Zaehle, S., and Papale, D.: Assessing and improving the representativeness of monitoring networks: The European flux tower network example, J. Geophys. Res., 116, G00J04, doi:10.1029/2010jg001562, 2011.

Taylor, G. I.: Eddy motion in the atmosphere, Philosophical Transactions of the Royal Society of London, A 215, 1-26, doi:10.1098/rsta.1915.0001, 1915.

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

Vaughan, A. R., Lee, J. D., Shaw, M. D., Misztal, P. K., Metzger, S., Vieno, M., Davison, B., Karl, T. G., Carpenter, L. J., Lewis, A. C., Purvis, R. M., Goldstein, A. H., and Hewitt, C. N.: VOC emission rates over London and South East England obtained by airborne eddy covariance, Faraday Discuss., 200, 599-620, doi:10.1039/c7fd00002b, 2017.

Villarreal, S., Guevara, M., Alcaraz-Segura, D., and Vargas, R.: Optimizing an environmental observatory network design using publicly available data, Journal of Geophysical Research: Biogeosciences, 124, 1812-1826, doi:10.1029/2018jg004714, 2019.

Wicker, L. J., and Skamarock, W. C.: Time-Splitting Methods for Elastic Models Using Forward Time Schemes, Monthly Weather Review, 130, 2088-2097, doi:10.1175/1520-0493(2002)130<2088:tsmfem>2.0.co;2, 2002.

Wiens, R. C., Neugebauer, M., Reisenfeld, D. B., Moses, R. W., Nordholt, J. E., and Burnett, D. S.: Genesis solar wind concentrator: Computer simulations of performance under solar wind conditions, Space Sci. Rev., 105, 601-626, doi:10.1023/a:1024474028352, 2003.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y. P.: Improving land surface models with FLUXNET data, Biogeosciences, 6, 1341-1359, doi:10.5194/bg-6-1341-2009, 2009.

Williamson, J. H.: Low-storage Runge-Kutta schemes, J. Comput. Phys., 35, 48-56, doi:10.1016/0021-9991(80)90033-9, 1980.

Wolf, B., Chwala, C., Fersch, B., Garvelmann, J., Junkermann, W., Zeeman, M. J., Angerer, A., Adler, B., Beck, C., Brosy, C., Brugger, P., Emeis, S., Dannenmann, M., De Roo, F., Diaz-Pines, E., Haas, E., Hagen, M., Hajnsek, I., Jacobeit, J., Jagdhuber, T., Kalthoff, N., Kiese, R., Kunstmann, H., Kosak, O., Krieg, R., Malchow, C., Mauder, M., Merz, R., Notarnicola, C., Philipp, A., Reif, W., Reineke, S., Rödiger, T., Ruehr, N., Schäfer, K., Schrön, M., Senatore, A., Shupe, H., Völksch, I., Wanninger, C., Zacharias, S., and Schmid, H. P.: The SCALEX campaign: Scale-crossing land surface and boundary layer processes in the TERENO-preAlpine Observatory, Bull. Am. Meteorol. Soc., 98, 1217-1234, doi:10.1175/bams-d-15-00277.1, 2017.

Wu, D., Wang, Z., Wechsler, P., Mahon, N., Deng, M., Glover, B., Burkhart, M., Kuestner, W., and Heesen, B.: Airborne compact rotational Raman lidar for temperature measurement, Opt. Express, 24, A1210-A1223, doi:10.1364/oe.24.0a1210, 2016.

Wu, P., Shen, H., Ai, T., and Liu, Y.: Land-surface temperature retrieval at high spatial and temporal resolutions based on multi-sensor fusion, Int. J. Digital Earth, 6, 113-133, doi:10.1080/17538947.2013.783131, 2013.

Wulfmeyer, V., Turner, D. D., Baker, B., Banta, R., Behrendt, A., Bonin, T., Brewer, W. A., Buban, M., Choukulkar, A., Dumas, R., Hardesty, R. M., Heus, T., Ingwersen, J., Lange, D., Lee, T. R., Metzendorf, S., Muppa, S. K., Meyers, T., Newsom, R., Osman, M., Raasch, S., Santanello, J., Senff, C., SpäTh, F., Wagner, T., and Weckwerth, T.: A new research approach for observing and characterizing land-atmosphere feedback, Bull. Am. Meteorol. Soc., 99, 1639-1667, doi:10.1175/bams-d-17-0009.1, 2018.

Wyngaard, J. C., and Brost, R. A.: Top-down and bottom-up diffusion of a scalar in the convective boundary layer, Journal of the Atmospheric Sciences, 41, 102-112, doi:10.1175/1520-0469(1984)041<0102:tdabud>2.0.co;2, 1984.

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

Xu, K., Metzger, S., and Desai, A. R.: Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance, Agric. For. Meteorol., 255, 81-91, doi:10.1016/j.agrformet.2017.10.011, 2018.

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

Xu, Z., Ma, Y., Liu, S., Shi, W., and Wang, J.: Assessment of the energy balance closure under advective conditions and its impact using remote sensing data, J Appl Meteorol Clim, 56, 127-140, doi:10.1175/jamc-d-16-0096.1, 2016.

Younker, D.: Value Engineering: Analysis And Methodology, CRC Press, Boca Raton, USA, 350 pp., 2003.

Zhang, L., and Pu, Z.: 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, doi:10.1155/2010/743863, 2010.

Zhang, Y., Ye, X., Wang, S., He, X., Dong, L., Zhang, N., Wang, H., Wang, Z., Ma, Y., Wang, L., Chi, X., Ding, A., Yao, M., Li, Y., Li, Q., Zhang, L., and Xiao, Y.: 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, Atmos. Chem. Phys. Discuss., 2020, 1-16, doi:10.5194/acp-2020-1168, 2020.

Ziehn, T., Law, R. M., Rayner, P. J., and Roff, G.: Designing optimal greenhouse gas monitoring networks for Australia, Geosci. Instrum. Method. Data Syst., 5, 1-15, doi:10.5194/gi-5-1-2016, 2016.