



Trade-Offs in Flux Disaggregation: A Large-Eddy Simulation Study

Matthias Sührling¹ · Stefan Metzger^{2,3} · Ke Xu³ · Dave Durden² · Ankur Desai³

Received: 2 March 2018 / Accepted: 2 August 2018 / Published online: 25 August 2018
© Springer Nature B.V. 2018

Abstract

Airborne flux measurements allow us to quantify the surface–atmosphere exchange over heterogeneous land surfaces. While often applied to regional-scale fluxes, it is also possible to infer component fluxes emanating from different surface patches from the measurement via disaggregation strategies. Here, we emulate flux disaggregation strategies by conducting an ensemble of virtual flight measurements within a set of large-eddy simulations over idealized surface heterogeneities and under different flow regimes. The resulting patch surface fluxes are compared with the prescribed patch surface fluxes in the simulation. To calculate fluxes along the flight legs, we apply traditional eddy-covariance and space–frequency (wavelet) methods. We show that the patch fluxes are captured best with the space–frequency method, where the disaggregation error is almost invariant of the segment length. For the eddy-covariance method, however, the error strongly depends on the segment length, with largest random and systematic errors for shorter segments. Furthermore, we determine a trade-off between a permissible disaggregation error and a sufficient resolution of the heterogeneous surface signals. Among our set-ups, an optimal segment length is determined to be 3–4 km for the eddy-covariance method, while with the space–frequency method even shorter segment lengths of a few hundreds of metres can be chosen, which enables sufficient isolation of signals from surface patches and the resolution of small-scale surface heterogeneity.

Keywords Aircraft measurement · Eddy covariance · Flux disaggregation · Flux footprint · Wavelet decomposition

✉ Matthias Sührling
suehring@muk.uni-hannover.de

¹ Institut für Meteorologie und Klimatologie, Leibniz Universität Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

² National Ecological Observatory Network, Battelle, 1685 38th Street, Boulder, CO 80301, USA

³ Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, 1225 West Dayton Street, Madison, WI 53706, USA

1 Introduction

The quantification of heat, moisture and trace gas exchanges between the Earth's surface and the atmosphere is crucial for improving modelling of the terrestrial ecosystem. Due to land-surface and microclimatic heterogeneities, the surface-atmosphere exchange can also be highly variable over a large range of spatial scales. Numerical weather prediction and climate models, however, are typically unable to resolve these heterogeneous surface-atmosphere interactions on scales smaller than the grid sizes (kilometre scale). Hence, heterogeneous surface-atmosphere interactions on the grid scale need to be parametrized. To develop and improve so-called flux-aggregation schemes in models, area-averaged observations of heat, moisture and trace gas fluxes are required, as well as information on the surface type that effectively contributes to the measured signal and how this signal is composed of source or sink contributions from different surface patches (Beyrich and Mengelkamp 2006). Airborne flux measurements are well suited to obtain information on both the regional as well as the local scale, and within a relatively short period of time (Desjardins et al. 1995), but methods to disaggregate this flux require further examination.

To date, no consensus has been reached regarding the appropriate strategy of how to disaggregate fluxes from area-averaged airborne measurements. Flight path segmentation based on land-use information (Samuelsson and Tjernström 1998; Vellinga et al. 2008) can be used to infer local surface fluxes from airborne eddy-covariance (EC) measurements. A similar more sophisticated approach for segmentation is the flux fragment method, which was originally developed by Kim et al. (2006) and further improved by Kirby et al. (2008) and Dobosy et al. (2017). The flux fragment method describes a conditional sampling method where a measured time series is divided into small fragments with a length of a few seconds, while for each fragment the footprint is calculated. If a fragment is dominated by a surface type, i.e. the source-area contribution from a surface type is higher than a certain threshold value, it is classified to a group of samples. Finally, surface fluxes are calculated from all classified fragments of the corresponding surface-type group. Even though segmentation methods have been already successfully applied, these approaches require a sufficient number of measurement points that are dominated by a specific surface type, i.e. the measurement height needs to be below the blending height in order to obtain homogeneous footprints where a single surface type dominates the signal. As the blending height is scale-dependent (Mahrt 2000), where blending over smaller heterogeneity scales occurs at lower levels compared to blending over larger scales, segmentation methods require aircraft flights at lower heights so as to capture the small-scale surface heterogeneity, which is, however, often not feasible due to safety considerations. Moreover, turbulent transport at scales larger than the fragment size is neglected, thus foregoing the unique ability of airborne flux measurements to provide unbiased estimates of the surface-atmosphere exchange (Foken et al. 2010).

An alternative strategy that overcomes the blending height issue has been proposed by Ogunjemiyo et al. (1998), who applied a quadratic regression model to infer local surface fluxes from measured fluxes based on eddy variance and corresponding source-area contributions of different surface types along the flight path. Hutjes et al. (2010) developed this strategy further and showed that a linear regression model is sufficient to disaggregate surface fluxes, i.e. that the measured flux is assumed to be a linear weighted average of surface fluxes emanating from different surface types. Further, by direct comparison of disaggregated fluxes against locally-measured surface fluxes, they revealed the practical potential of the disaggregation strategy. As with the flux fragment method,

a remaining critical issue is the choice of the averaging length for calculating the EC fluxes. For example, on the one hand, if the averaging length is too large, the spatial resolution of observed fluxes along the flight paths possibly averages out the surface heterogeneity signals, and disaggregation is no longer possible. On the other hand, if the averaging length is too small, larger scales that significantly contribute to the flux are filtered out, e.g. large eddies or heterogeneity-induced mesoscale circulations (Mauder et al. 2007).

To overcome these issues, Metzger et al. (2013) replaced the flux computation in the space domain with a space–frequency approach using a wavelet transformation. In this way, fluxes with high spatial resolution along the flight path can be calculated without neglecting flux contributions on larger scales. Even though all relevant scales are captured by this approach, it should be noted that the random error for each resulting flux observation is still a function of the spatial resolution. Thus, high-resolution observations (or small averaging lengths) still result in high random uncertainties (Lenschow et al. 1994). While these uncertainties enter the forenamed disaggregation methods, their net effect is essentially compensated through the ensemble effect of a proportionally larger sample size (Mahrt 1998; Metzger et al. 2013).

Here, given the range of approaches in the literature, we ask, what is the general performance of linear disaggregation for different scales and amplitudes of surface heterogeneity, as well as under different flow regimes? Further, what are the uncertainty bounds regarding averaging length? To answer these questions, we emulate the linear disaggregation strategy proposed by Hutjes et al. (2010) using virtual aircraft measurements performed in large-eddy simulations (LES), while flux calculations are based on the traditional EC method as well as a space–frequency method (see Metzger et al. 2013). For this purpose we use LES datasets based on idealized one-dimensional stripe-like surface heterogeneities. Previous studies of virtual aircraft flights using LES revealed systematic errors in measurement strategies, which could hardly be detected in ‘real-world’ measurements with unknown boundary conditions (Schröter et al. 2000; Sühling and Raasch 2013).

Another source of uncertainty affecting the disaggregation of surface fluxes is the calculation of the effective area of surface influence for each flux observation, i.e. the flux footprint. In order to determine the flux footprint, previous studies aggregated turbulence statistics along parts of the flight leg. These were then used in footprint parametrizations with far-reaching assumptions concerning boundary-layer structure and homogeneity of the upstream flow field. In order to minimize the uncertainty imposed by footprint modelling in the present study, we use a Lagrangian particle model already embedded into the LES model (Steinfeld et al. 2008) to determine flux footprints. Even though this approach is not free of uncertainty, the LES footprint calculation explicitly addresses the main assumptions underlying forenamed parametrizations, and no aggregation of turbulence statistics is required.

The paper is organized as follows: Sect. 2 describes the LES model and the embedded Lagrangian particle model, the footprint calculation from the LES, as well as the simulation set-ups. Furthermore, virtual measurements are described, as well as the flux calculations and the framework of the applied disaggregation. Section 3 first gives an overview of the boundary-layer structure, followed by results on the general performance and uncertainty of the disaggregation using EC and space–frequency methods. Finally, in Sect. 4, a summary and ideas for future studies are presented.

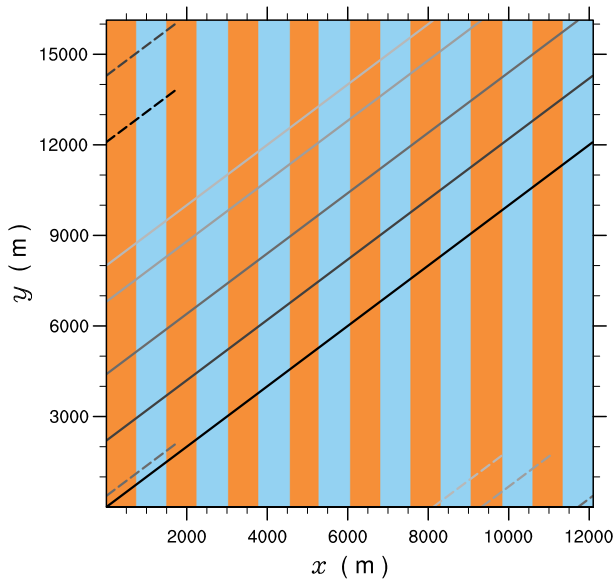


Fig. 1 Stripe-like surface pattern shown for simulations A50W075U. Orange (blue) coloured contours indicate warm–dry (cold–wet) patches. Black and grey coloured lines indicate the alignment of the five flight legs, while solid and dashed lines indicate the flight legs before and after cyclic conditions for the virtual aircraft apply for the first time, respectively. Note, for the sake of clarity, the entire legs are not shown

2 Methods

2.1 LES Model and Numerical Experiments

The Parallelized LES Model PALM (Raasch and Schröter 2001; Maronga et al. 2015), revision 1317¹, was used for the numerical simulations. PALM solves the non-hydrostatic incompressible Boussinesq equations, and for the subgrid model the kinetic energy scheme of Deardorff (1980) is used. The advection terms were discretized using a fifth-order scheme (Wicker and Skamarock 2002), and a third-order Runge–Kutta scheme by Williamson (1980) was used for the time integration. At the lateral boundaries, cyclic conditions were applied. As surface boundary condition for the momentum equations, Monin–Obukhov similarity theory (MOST) was applied locally between the surface and the first vertical grid level, while for the surface sensible and latent heat fluxes values constant in time were prescribed. At the beginning of each simulation, randomly distributed perturbations were imposed on the horizontal velocity fields to initiate turbulence.

All numerical experiments were initialized with a potential temperature $\theta(z)$ and specific humidity profile $q(z)$, which had constant values up to 800 m. This layer was capped by a dry inversion with a temperature (humidity) jump of 2 K ($-0.0005 \text{ kg kg}^{-1}$), and the free atmosphere above had a temperature lapse rate of 0.006 K m^{-1} with constant humidity. A homogeneous roughness length of $z_0 = 0.1 \text{ m}$ was used. We divided the model surface into stripe-wise alternating patches of higher sensible heat flux H_0 (lower latent heat flux L_0) and lower H_0 (higher L_0) along the x -axis, respectively, illustrated in Fig. 1, hereafter referred to as warm–dry and cold–wet patches, respectively. Along the y -axis, the model

¹ The code can be accessed under <https://palm.muk.uni-hannover.de/trac/browser/palm?rev=1317>.

Table 1 Parameters of the performed simulations

Name	Patch width (m)	$H_{0,wd} \mid H_{0,cw}$ (W m^{-2})	$L_{0,wd} \mid L_{0,cw}$ (W m^{-2})	Bowen Ratio warm cold	U_{BL} m s^{-1}	V_{BL} m s^{-1}
A50W300U	3000.0	150 50	150 250	1.0 0.2	5.0	0.0
A50W300V	3000.0	150 50	150 250	1.0 0.2	0.0	5.0
A50W150U	1500.0	150 50	150 250	1.0 0.2	5.0	0.0
A50W150V	1500.0	150 50	150 250	1.0 0.2	0.0	5.0
A50W075U	750.0	150 50	150 250	1.0 0.2	5.0	0.0
A50W075V	750.0	150 50	150 250	1.0 0.2	0.0	5.0
A20W300U	3000.0	120 80	180 220	0.67 0.36	5.0	0.0

$H_{0,wd}$ and $H_{0,cw}$ indicate the prescribed surface sensible heat flux at the warm–dry and cold–wet patches, respectively (equivalent for the surface latent heat flux L_0). U_{BL} and V_{BL} are the horizontal mean velocity components along and perpendicular to the heterogeneous x -axis within the boundary layer during the simulation, respectively. Please note, in order to better relate the prescribed fluxes to reality, flux values are given in units of W m^{-2} , assuming an air density of 1 kg m^{-3} .

surface was homogeneous. The domain-averaged surface fluxes were constant in time with values of $H_0 = 0.1 \text{ K m s}^{-1}$ and $L_0 = 8 \times 10^{-5} \text{ m s}^{-1}$, which correspond to heat fluxes of $\approx 100 \text{ W m}^{-2}$ and $\approx 200 \text{ W m}^{-2}$ (assuming a constant air density of 1 kg m^{-3}), respectively. In order to create different heterogeneous forcing, we varied the width of the surface patches among the simulations, with patch widths W of 0.75 km, 1.5 km, and 3 km. In addition, for a fixed patch width we also varied the amplitude of the surface heterogeneity.

In our simulations, the geostrophic wind speed was adjusted such that the mean wind speed within the boundary layer was 5 m s^{-1} . Previous studies have shown that the impact of surface heterogeneity on the boundary-layer structure depends strongly on the mean flow direction (Raasch and Harbusch 2001). Hence, we decided to create two sets of heterogeneous simulations, one with flow parallel to the y -axis, where an air parcel experiences the same surface forcing over a longer period of time and mesoscale circulations can develop, and one set-up where the flow is parallel to the x -axis and the development of mesoscale circulations is suppressed (see Sühling et al. 2014).

The model domain was $12 \text{ km} \times 16 \text{ km} \times 3.5 \text{ km}$ in the x -, y -, and z -directions, respectively, and an isotropic grid spacing of 7 m was used. A sensitivity study (not shown) indicated this grid spacing to be sufficient, as turbulence and LES footprint statistics at the lowest measurement level, which are quite sensitive to grid resolution (Steinfeld et al. 2008), did not change for further decreases in grid spacing. The simulations ran for 5 h of simulation time, while data analysis started after 2 h. During the data analysis period the timestep was kept constant at a value of 0.3 s, in order to perform high-resolution virtual measurements. An overview of all performed simulations is given in Table 1. The name of the particular simulation is derived from the respective amplitude and patch width, as well as the prescribed flow direction using the u -component (v -component) of velocity indicating flow parallel to the x -axis (y -axis). For example, A50W300U is a simulation with flow parallel to the x -axis, as well as a patch width of 3 km and an amplitude of 50% with respect to the horizontal mean, where $H_0(x)$ is 50% higher (lower) over the warm–dry (cold–wet) patch compared to the horizontal mean, respectively.

2.2 Embedded Lagrangian Particle Model and Footprint Evaluation

For the flux-footprint evaluation, a Lagrangian particle model embedded into the LES model was used, according to Steinfeld et al. (2008) and Markkanen et al. (2009, 2010). This approach, based on Weil et al. (2004, 2012), separates the particle speed into a deterministic and a stochastic part, which corresponds to dividing the turbulent flow field into a resolved-scale and a subgrid-scale (SGS) portion, respectively, following the LES philosophy. The resolved-scale velocity is provided by the LES at each timestep, while the SGS velocity is predicted by integrating a stochastic differential equation according to Weil et al. (2004), who strictly adopted the Thomson (1987) model to the subgrid scale by assuming isotropic and Gaussian-distributed turbulence. To parametrize the stochastic particle dispersion on the subgrid scale, the LES provides local values of the SGS turbulent kinetic energy and the dissipation rate at each timestep. According to Steinfeld et al. (2008), the LES data are interpolated bi-linearly on the actual particle position in the horizontal. In the vertical a linear interpolation is used, except for the particles located between the surface and the first grid level, where a logarithmic interpolation according to local MOST for the resolved-scale horizontal velocity components is applied. At the bottom boundary of the model domain we used a reflection boundary condition for the particles, and cyclic conditions at the lateral boundaries. A more detailed description of the particle model embedded into the LES model is given by Maronga et al. (2015).

Flux footprints $f(x, y)$ were calculated by a post-processor according to Kurbanmuradov et al. (2001),

$$f(x, y) = \frac{1}{N_{\text{eff}} \Delta x \Delta y} \sum_{i=1}^{N_p} \sum_{j=1}^{n_i} \frac{w_{i,j}^p}{|w_{i,j}^p|} I(x - X_{i,j}, y - Y_{i,j}), \quad (1)$$

where N_p is the total number of particles, n_i is the number of intersections of the trajectory of particle i with the sensor at the measurement level, $w_{i,j}^p$ is the vertical component of the particle speed at the respective intersection point with the sensor, and $X_{i,j}$, $Y_{i,j}$ denote the horizontal distance between the source coordinates of particle i at the j -th intersection with the sensor in the x - and y -directions. Since the probability that a particle passes through a sensor at a certain point in space is rather low, we used a horizontally extended sensor area with edge lengths of 7 m, which should be fine as long as the sensor area is small compared to the underlying surface heterogeneity. Note that N_{eff} is the effective number of particles that is used for normalization, i.e. the number of particles released from an area equal to the sensor size, and with a uniform particle source as used herein, N_{eff} is identical for each location, while for heterogeneous particle sources N_{eff} is a function of x, y (Steinfeld et al. 2008). Further, $\Delta x = \Delta y = 7$ m are the grid spacings of the analysis grid in the post-processor, and $I = 1$ where the particle trajectory intersects the given measurement level in the coordinate interval $-\Delta x/2 \leq X_{i,j} < \Delta x/2$ and $-\Delta y/2 \leq Y_{i,j} < \Delta y/2$, else $I = 0$.

In order to obtain more contiguous and statistically more reliable flux footprints at each discrete x -position, we averaged $f(x, y)$ over the homogeneous y -direction. Thus, we follow Markkanen et al. (2010), who pointed out that for a similar heterogeneous set-up footprint statistics for a given $x - y$ position are not sufficient since too few particles cross the virtual sensor area.

2.3 Virtual Measurements

In order to mimic ‘real-world’ aircraft measurements, we performed temporally high-resolution virtual aircraft measurements according to Schröter et al. (2000) and Sühling and Raasch (2013), where a virtual sensor moves through the model domain at a given speed of 25 m s^{-1} and samples space–time series of θ , q , and vertical velocity w at each LES timestep. As the sensor position does not necessarily coincide with the numerical grid, data are bi-linearly interpolated onto the sensor position. Virtual measurements were taken between 2 and 5 h of simulation time at a sampling rate of 3.3 s^{-1} (corresponding to the LES timestep). The measurement height was 49 m, where the simulated flow is well resolved and the contribution of SGS fluxes to the vertical transport is less than 1%, but the measurement level is still in the surface layer where the fluxes are constant within 5–10% such that the measured fluxes can be interpreted as surface fluxes. Note aircraft are rarely able to fly over most surfaces at heights of 49 m, but instead they fly multiple higher levels in the boundary layer to infer surface fluxes by interpolation of a linear flux profile, assuming constant vertical flux divergence. For simplicity, we avoided this additional calculation and used the 49-m level instead.

Five diagonal legs are flown parallel to each other, indicated in Fig. 1, which are well separated in space to achieve statistically independent measurements. Similar to the LES flow, cyclic conditions are assumed for the virtual aircraft flights, i.e. when an aircraft leaves the model domain, it enters it again at the opposite side, which is illustrated in Fig. 1 by the solid and dashed lines.

2.4 Flux Calculation

The legs were divided into equally-spaced, non-overlapping segments of length l , ranging from 200 m up to 12.5 km. Turbulent vertical sensible and latent heat fluxes were calculated from the virtually sampled space series for each segment, using the classical time–domain EC method, as well as space–frequency wavelet decomposition. With respect to the classical EC method, turbulent fluxes were calculated by averaging the product of the turbulent deviations of w and θ (or q , respectively) over the respective segment. Turbulent deviations were computed by linear de-trending of the respective space series over each segment. One of the main shortcomings in doing this, particularly for shorter segments, is the removal of scales larger than the respective segment contributing to the flux, e.g. by thermal convection or by mesoscale motions (Finnigan et al. 2003). In order to overcome this, continuous wavelet transformation is used (Strunin and Hiyama 2005; Mauder et al. 2007; Metzger et al. 2013), which allows study of the covariance of signals not only in the space domain, but also in the frequency domain, so that larger-scale contributions to the flux can also be considered. Following Torrence and Compo (1998), we applied a continuous wavelet transform approximation. A continuous wavelet transform of a sequence of data $r(n)$ is the convolution of $r(n)$ with a wavelet function

$$W_r(a, b) = \sum_{n=0}^N r(n) \psi_{a,b}^*(n), \quad (2)$$

where $W_r(a, b)$ are the wavelet coefficients, where a is a parameter representing different discrete scales in the frequency domain and b is the location in the space (or time) domain; N is the number of discrete data points of the entire sampled series; $r = (n - b)/a$ is a

dimensionless coordinate in the frequency–space domain, and ψ^* is the complex conjugate (denoted by the asterisk) of the wavelet function

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi \left(\frac{n-b}{a} \right), \quad (3)$$

where $\psi \left(\frac{n-b}{a} \right)$ is called mother wavelet. Following Metzger et al. (2013), we chose the Morlet wavelet $\psi(r) = \pi^{-1/4} \exp(i\omega_0 r) \exp(-r^2/2)$ as the mother wavelet, with $\omega_0 = 6$, which has been proven to be well suited for analyzing turbulence data (Strunin et al. 2004; Thomas and Foken 2005).

The covariance $COV_{r,s}$ of two discrete data series r and s , is defined as

$$COV_{r,s} = \frac{\delta j \delta t}{C_\delta N} \sum_{j=0}^J \sum_{n=0}^{N-1} \frac{W_r(a_j, b_n) W_s^*(a_j, b_n)}{a_j}, \quad (4)$$

where $C_\delta = 0.776$ is a characteristic constant for the Morlet wavelet, and δj is a non-dimensional factor determining the spacing between the scales in the frequency space.

The resulting wavelet transformation is affected by edge effects, also known as the cone of influence, which tapers towards the centre of the dataset (Torrence and Compo 1998). However, calculating covariances using frequency scales below the cone of influence implies less reliable estimates near the edges and thus biased fluxes. In fact, when we used the entire space series for the flux computation, including scales affected by the cone of influence, we experienced a large bias in the resulting wavelet fluxes (not shown). Hence, we applied a rectangular-shape cut-off filter to avoid effects due to the cone of influence in the resulting fluxes, which yielded consistent flux integration scales along the flight path and a convergence of the results. Therefore, as a first step, we determined a lower frequency scale at which all eddies are captured sufficiently. An ogive analysis (see also Sect. 3.3) showed that contributions at frequencies smaller than 10^{-2} s^{-1} are no longer significant. Considering a safety factor, we finally set the lower frequency scale to $1.6 \times 10^{-3} \text{ s}^{-1}$ (assuming Taylor's hypothesis this would correspond to a spatial scale of 15 km) in our case. In the following, this lower frequency scale was used to cut-off the edges of the space series used for the flux calculation, by flagging the positions where the lower frequency scale intersects the cone of influence. Hence, fluxes were calculated according to Eq. 4, except for cutting-off the leading and trailing edges of the space–time series and setting a lower integration bound in frequency space, so that fluxes are no longer biased due to cone of influence effects.

2.5 Flux Disaggregation

We follow the approach proposed by Hutjes et al. (2010) to disaggregate a measured flux into its component fluxes emanating from different surface types, by applying a tile approach (Dolman and Blyth 1997; Michaud and Shuttleworth 1997), where the observed segment flux F_{obs} over a heterogeneous landscape is the area-weighted sum of all the component fluxes $F_{i,0}$ within the footprint of the measurement,

$$F_{\text{obs}} = \sum_{i=1}^{N_{\text{het}}} A'_i F_{i,0}, \quad (5)$$

with A'_i being the relative source-area contribution of the i -th surface type over a single segment. To obtain the relative contributions A'_i from the warm and cold patches, we made use

of the calculated footprints in our simulations, i.e. for each discrete x -position we integrated the footprint f over the respective area of surface type i and averaged the resulting local contributions over the segment length.

For the sake of clarity, in our cases the observed fluxes are either provided by the classical EC method or the wavelet method. Finally, merging all observations and source-area contribution gives a linear equation system

$$\mathbf{F}_{\text{obs}} = A\mathbf{F}_0, \quad (6)$$

with $\mathbf{F}_{\text{obs}}$ being a p -by-1 vector of observed segment fluxes (eddy covariance or wavelet), and p being the number of non-overlapping segments. $\mathbf{F}_0$ is a N_{het} -by-1 vector of regression coefficients (disaggregated fluxes), and A is a p -by- N_{het} matrix of relative source-area contributions for each segment, with $N_{\text{het}} = 2$ being the number of surface types.

The individual surface fluxes $\mathbf{F}_0$ emanating from the different patches can now be obtained by solving Eq. 6 using a robust linear regression method (we used the ‘rlm’ algorithm with bi-square estimating function from the R ‘MASS’ package) with zero regression intercept, according to Hutjes et al. (2010).

At this point we note that at the measurement level $z = 49$ m the domain-averaged flux is already about 5% lower compared to its surface value due to vertical flux divergence (see Fig. 3), indicating that also perfectly disaggregated fluxes are likely to show about a 5% offset compared to the ‘true’ patch fluxes. To overcome this in the ‘real world’, measurements should be taken at multiple heights to estimate the vertical flux divergence above the warm and cold patches and correct the disaggregated fluxes appropriately. However, insufficient flux sampling or spatially different vertical divergences due to mesoscale circulations might introduce uncertainty in the results. Hence, we refrained from doing this and simply down-adjusted the disaggregated fluxes to the surface by adding the difference between the domain-averaged flux at $z = 49$ m and its domain-averaged value at the surface, resulting in a correction of about 5%, constant in time, both sensible and latent heat fluxes.

3 Results

3.1 Boundary-Layer Structure

Figure 2 shows vertical profiles of domain- and phase-averaged potential temperature and specific humidity. As expected, the domain-averaged profiles show a vertically well-mixed convective boundary layer with a dry capping inversion. For the sake of clarity, phase-averaged profiles over the centre of the cold and warm patches are only shown for the simulations with largest patch width. For simulation A50W300V, where the flow is parallel to the surface heterogeneity, the boundary layer is warmer and drier over the warm patch throughout the bulk of the boundary layer, indicating that the boundary layer is not well mixed horizontally, as already observed in previous studies with similar set-ups (Ouwensloot et al. 2011). In contrast, for simulation A50W300U, where the flow is perpendicular to the stripe-like surface heterogeneity leading to an increased horizontal mixing, θ and q over the cold and warm patches are similar throughout the depth of the boundary layer, except for the near-surface region, which is a bit warmer and drier over the warm-dry patch compared to the cold-wet patch.

Figure 3 shows vertical profiles of the sensible and latent heat fluxes. For all simulated cases the profiles decrease linearly with height within the boundary layer, indicating a ver-

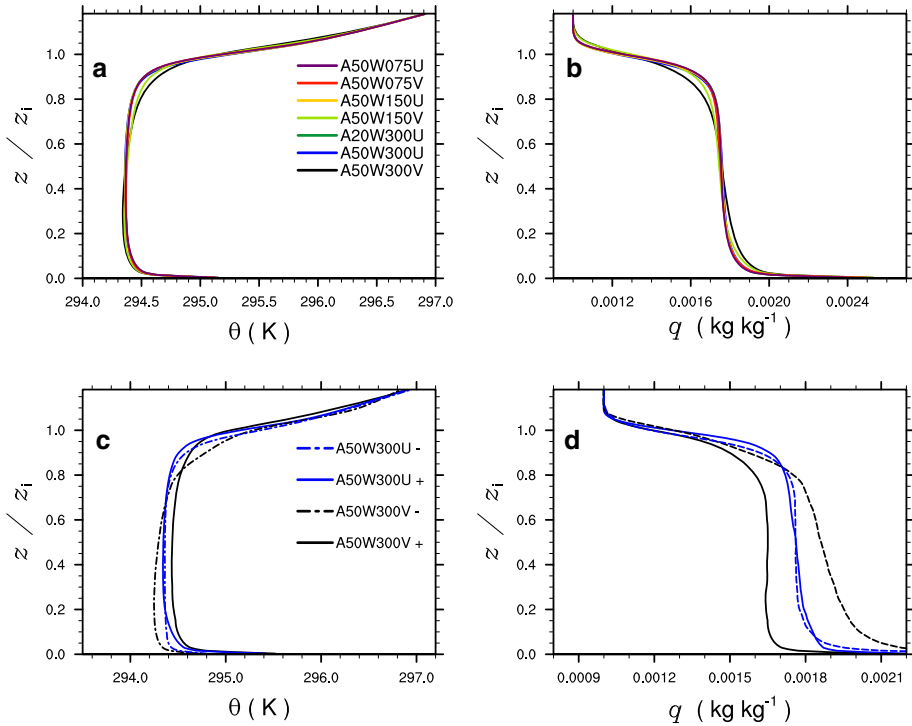


Fig. 2 Vertical profiles of domain-averaged (a, b) and phase-averaged (c, d) θ and q after 3 h of simulation time. The vertical coordinate is normalized with the boundary-layer depth z_i , which is at ≈ 1200 m for all simulations. Domain-averaged profiles are shown for all simulations, while for the sake of clarity, phase-averaged profiles are only shown for the simulations A50W300U and A50W300V. The *plus* and *minus* symbols indicate profiles over the centre of the warm and the cold patches, respectively. All profiles were time-averaged over 30 min

tically well-mixed state. The profiles are almost the same among the heterogeneous cases, except for the top of the boundary layer, where e.g., simulations A50W150V and A50W300V show larger negative values that can be attributed to increased entrainment due to the heterogeneous surface forcing (Sührling et al. 2014), leading to slightly different vertical flux divergences within the boundary layer. Moreover, the latent heat flux at the boundary-layer top is significantly smaller than the surface flux, indicating that the moisture in the boundary layer is mostly surface-driven.

Figure 4 shows $x - y$ cross-sections for the time-averaged vertical velocity at a height of 49 m where the virtual measurements took place. In the case of simulation A50W300V, mean updraft and downdraft motions are located over the warm and cold patches, respectively, indicating well-developed heterogeneity-induced mesoscale circulations that extend up to the top of the boundary layer (not shown). For a similar set-up, e.g. van Heerwaarden and Vila-Guerau de Arellano (2008) have shown that these circulations contribute significantly to the vertical transport of heat and moisture, with the largest contributions up to 10% (40%) for the sensible (latent) heat flux in the middle part of the boundary layer to the total transport. Even though the circulations are usually weaker near the surface, the corresponding mean vertical motion is still on the order of $0.1 - 0.2 \text{ m s}^{-1}$ at the measurement level. In the case of simulation A50W300U, no mesoscale circulation can be observed, which is

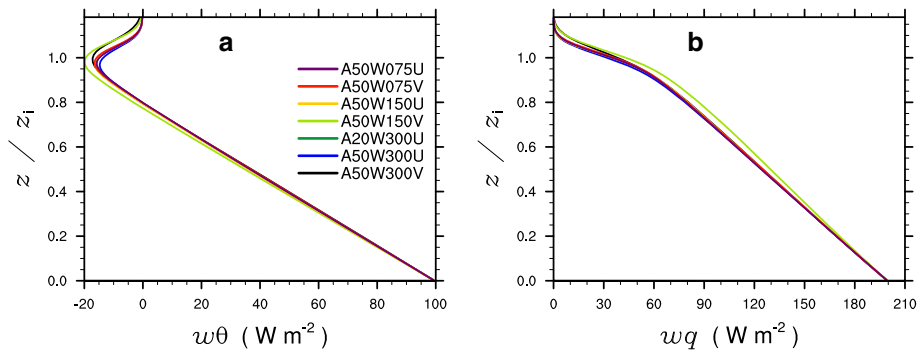


Fig. 3 Domain-averaged vertical profiles of, **a** sensible heat flux, and **b** latent heat flux, averaged over 30 min after 3 h of simulation time. The vertical z coordinate is normalized with the boundary-layer depth

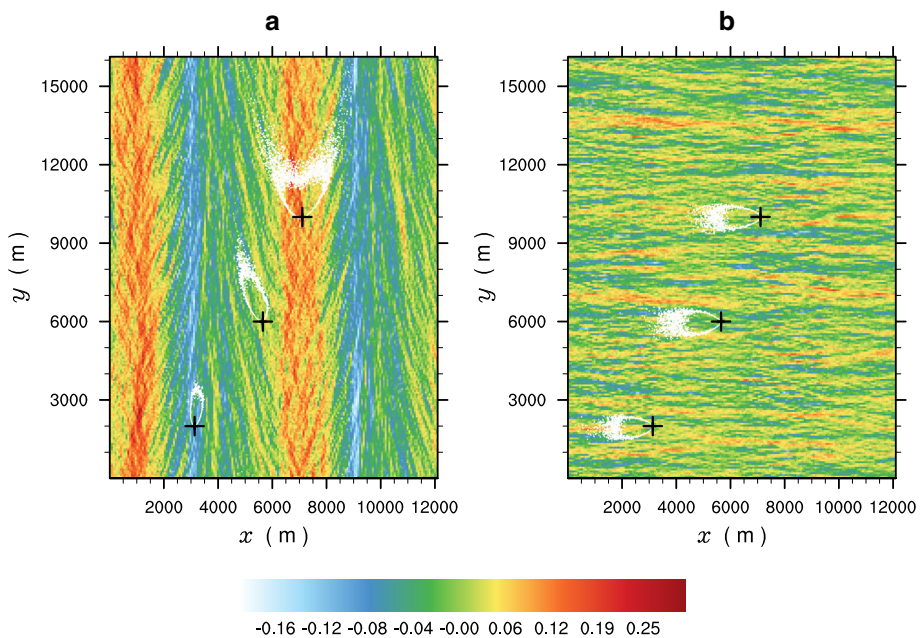


Fig. 4 x - y cross-sections of w (m s^{-1}) after 3 h of simulation for **a** simulation A50WL300V, and **b** A50WL300U at a height of $z = 49$ m, averaged over 30 min. White contour lines show the 90% cumulative flux-footprint extent at different virtual sensor positions above the centre of the cold patch (lower-left sensor), at the transition between cold and warm patches (central sensor), and above the centre of the warm patch (upper-right sensor), all at a height of $z = 49$ m. Please note the difference in flow directions between simulation A50WL300V and A50WL300U, and thus the different footprint orientation

attributed to the increased horizontal mixing by the mean flow perpendicular to the surface heterogeneities, preventing the build-up of horizontal temperature gradients and thus the development of mesoscale circulations. To note, according to previous findings by Raasch and Harbusch (2001) for similar stripe-like surface heterogeneity, the other parallel-flow simulations (A50W150V and A50W075V) also exhibit mesoscale circulations of similar strength, while the perpendicular-flow simulations show no circulations at all (not shown).

3.2 Footprints

In order to visualize the footprint area at a given sensor position, we estimated the smallest source area that contributes to 90% of the signal at that sensor. For different sensor positions, Fig. 4 shows the respective 90% footprint extent. Close to the sensor (1–2 km upwind), where the bulk of the surface signal originates from (not shown), the 90% footprint appears continuous. Farther upwind of the sensor, however, the 90% footprint area is more scattered and non-continuous, which appears as white dots in Fig. 4 and can be attributed to an insufficient number of particles (originating from these regions) passing the sensor. Since the relative contribution from these locations farther upwind to the measured signal at the sensor is rather low, however, the impact on further results is expected to be negligible.

In the case of simulation A50WL300U, the footprints at the different sensor locations look almost similar, and no significant differences concerning size and shape can be observed. A completely different situation can be observed in simulation A50WL300V, where the respective footprints differ significantly among the sensor locations. Over the cold patch the 90% area is significantly smaller, particularly in the crosswind direction, compared to the other locations. This can be attributed to the weaker surface heating over the cold patch leading to less convective updrafts, as well as to the occurrence of the subsiding branch of the mesoscale circulation, which diverges horizontally close to the surface, further weakening convective updrafts (Steinfeld et al. 2008; Markkanen et al. 2010). In contrast, over the centre of the warm patch the footprint is significantly larger, due to the increased surface heating. It is emphasized that the footprint exhibits two branches with tails extending towards the north-west and north-east, which is due to the convergence of the mesoscale circulation over the centre of the warm patch (Markkanen et al. 2010). For the sensor location between the cold and warm patches, the footprint mainly arises from the cold patch, further illustrating the impact of the mesoscale circulation.

Figure 5 shows the relative source-area contributions along the x -axis for the warm and cold patches, where the contributions of the warm and cold patches sum to 1 at each location. The shape differs between the different flow set-ups. For the parallel-flow simulations (see Fig. 5a–c) the local source-area contributions show symmetric oscillation with an abrupt increase at the patch borders of the relative contribution of the respective patch. In particular for the largest surface heterogeneities, the measured signals originate almost entirely from the underlying patch, i.e. the footprint is homogeneous, which is most obvious for the cold patch, whereas for sensors above the edge regions of the warm patch low contributions from the cold patch can still be observed. By contrast, for the perpendicular-flow simulations (see Fig. 5d–g), the contributions are not symmetric but exhibit a fin-like shape, with a smoother increase of the corresponding patch contribution to the total signal, indicating more mixed source-area contributions.

3.3 Ogive Analysis

As a next step, we estimated the temporal/spatial scales that contribute to the measured fluxes, which is mainly motivated by two reasons. First, we performed virtual measurements at a sampling rate of 3.3 s^{-1} , which is about one-half to one order of magnitude lower compared to physical measurements in the ‘real world’. This gives rise to the question of whether the turbulence spectrum is sufficiently well captured to avoid significant bias due to high-frequency flux loss. Secondly, as mentioned above, we also applied the EC method to calculate fluxes, which is known to miss larger-scale low-frequency contributions if the

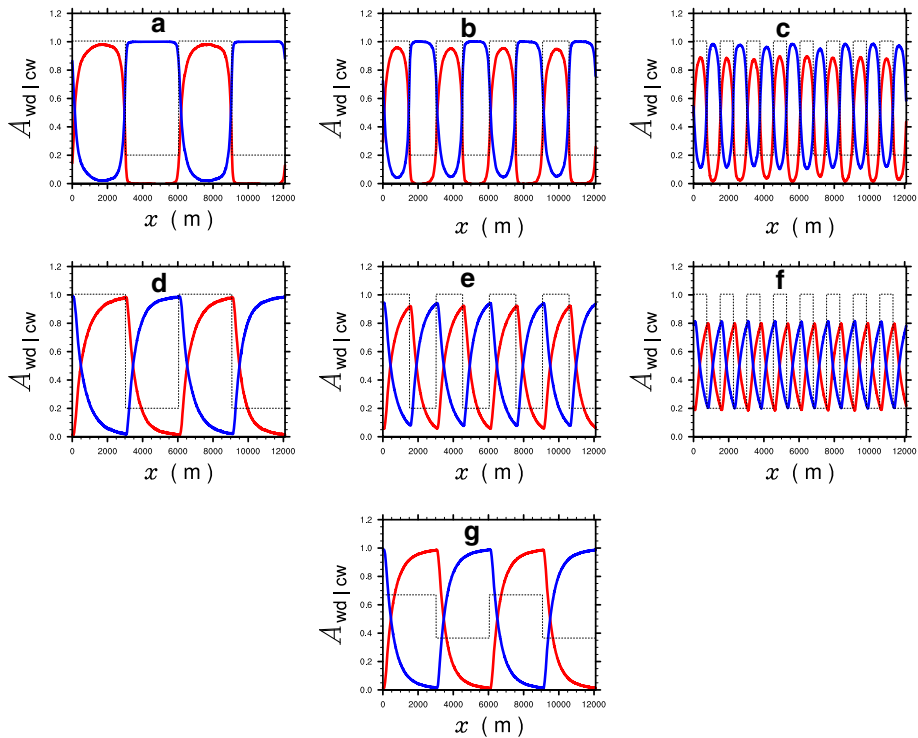


Fig. 5 Relative source-area contributions at $z = 49$ m of the warm-dry (wd) patch (red) and the cold-wet (cw) patch (blue) along the heterogeneous x -axis, for **a** A50WL300V, **b** A50WL150V, **c** A50WL075V, **d** A50WL300U, **e** A50WL150U, **f** A50WL075U and **g** A20L3U. The box-like dashed line indicates the location of the warm-dry and cold-wet patches

segment length is too short. This prompts the question: which segment length is long enough to capture all relevant scales with the EC method? In order to estimate which scales contribute to the flux, we used the ogive function, proposed by Oncley et al. (1990),

$$O_{w\phi}(f) = \int_{f_0}^{\infty} COV_{w\phi}(f) df, \quad (7)$$

which is the cumulative integral of the covariance spectrum between w and $\phi \in (\theta, q)$, with frequency f ; f_0 is the Nyquist frequency, i.e. half the sampling rate of the measurement. Figure 6 shows the ogive functions for the sensible and latent heat flux, derived from the virtual measurements. Contributions at high frequencies close to the Nyquist frequency are small, indicating that further increasing the sampling rate would not significantly increase the flux. We note that this argumentation by itself is insufficient to prove negligible high-frequency loss, because the measurement itself can only capture those parts of the turbulent spectrum that are resolved by the numerical grid in the LES model. Nevertheless, due to the fact that the unresolved SGS motions contribute less than 1% to the total flux at the measurement level, we are confident that the applied sampling rate and flow resolution are sufficient.

On the larger scales, the ogives for the sensible heat flux are very similar, except for simulations A50WL300V and A50WL150V, where the ogives are smaller in the spectral

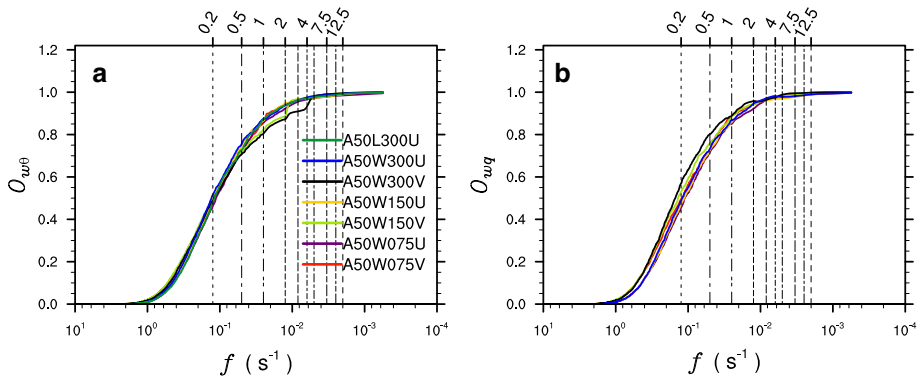


Fig. 6 Ogive functions for, **a** sensible heat flux, and **b** latent heat flux for all simulations. For the sake of comparison, ogive functions are normalized to 1. The dashed vertical lines indicate the corresponding segment lengths (in km), labelled above

range between 10^{-1} and 10^{-2} s^{-1} (a few hundreds of metres to a few kilometres), which coincides well with the dimensions of the existing mesoscale circulations in these cases. At frequencies smaller than 10^{-2} s^{-1} , which corresponds to segment lengths larger than 2.5 km, all ogives approach one, indicating that all relevant scales beyond are captured. Viewed the other way around, EC fluxes calculated on shorter segments suffer significantly from missing larger-scale contributions.

3.4 Which of the Approaches Best Captures the Patch Fluxes?

Figure 7 shows disaggregated fluxes based on EC- and wavelet-based segment fluxes for simulation A50WL300V and all considered segment lengths against the prescribed LES fluxes. The disaggregated fluxes based on the EC method show a wide spread among the different segment lengths, with largest deviations from the prescribed fluxes for shorter segment lengths, significantly underestimating the ‘true’ patch fluxes. For the shorter segments, the EC-based fluxes in part depict nearly homogeneous surfaces fluxes, e.g. for $l = 200 \text{ m}$ or 500 m . With increasing segment length, the sampling of larger scales contributing to the flux improves for the EC method, and the EC-based disaggregated fluxes approach the ‘true’ patch fluxes, even though the ‘true’ patch fluxes are still underestimated. For the wavelet method, the disaggregated cold-patch fluxes are slightly overestimated for sensible and latent heat, while the warm-patch fluxes agree fairly well with the respective ‘true’ patch flux. In contrast to the EC method, the disaggregated fluxes agree well with the ‘true’ fluxes, independent of the applied averaging length.

3.5 Uncertainty of Observed Segment Fluxes

In order to investigate the general offset between disaggregated EC- and wavelet fluxes, we were further investigating the uncertainty of the measured segment fluxes. Based on integral length scales, Lenschow et al. (1994) proposed estimates to assess the statistical uncertainty of a measured flux, which depend primarily on the applied averaging length. However, for relatively small averaging lengths, these error estimations can significantly underestimate the actual uncertainty (Sührling and Raasch 2013). Hence, we decided to evaluate the random

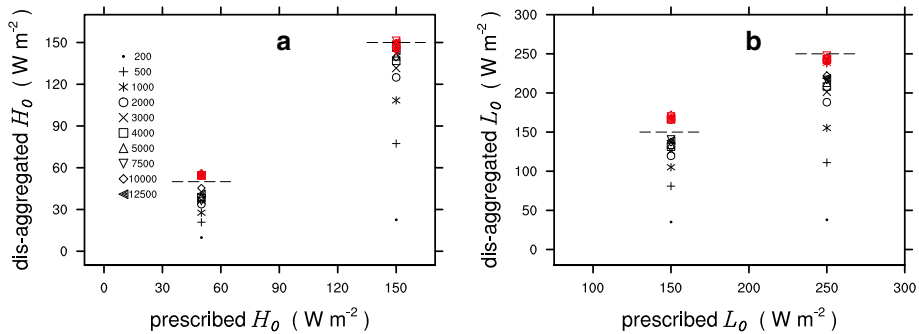


Fig. 7 Disaggregated fluxes for, **a** sensible heat flux, and **b** latent heat for simulation A50WL300V for different segment lengths, indicated by the markers. Black-coloured markers indicate EC-based fluxes, red-coloured wavelet-based fluxes. The horizontal dashed lines indicate ‘true’ patch fluxes

uncertainty by making use of the relative standard deviation σ of all measured fluxes for segments with similar source-area contributions. Therefore, we separated the source-area composition observed at all segments into bins of equal size; afterwards, we calculated σ over all fluxes observed at segments with similar source-area composition. Figure 8a,b shows the relative standard deviation of the measured fluxes according to their source-area composition for different segment lengths for simulation A50WL300V. First, it is obvious that for increasing segment lengths the range of variations in source-area composition between warm and cold patches becomes smaller. For instance, for $l = 1000$ m, the contributions from the warm and cold patches range from $\approx 20 - 80\%$, while for $l = 10,000$ m the contributions only range from $\approx 42 - 58\%$, which becomes evident if one applies a running average over the source-area contributions (see Fig. 5) with given segment length. For further increasing segment lengths, the variations in the source area would tend towards equal compositions, being the limit for solving Eq. 5 as A becomes linearly dependent and a unique solution no longer exists, hence, in terms of linear regression, shorter segment lengths are preferable to also sample very small-scale surface heterogeneity. Moreover, long segment lengths will be also detrimental to the ability to scale up to an entire ‘real-world’ domain and cover the corresponding wide range of surface properties appropriately.

For the sensible heat flux, Fig. 8a, b indicates an increasing σ for increasing contributions from the warm-dry patch, which could be attributed to stronger convection, so that the measured flux becomes more prone to random sampling of stronger updrafts and downdrafts. For the latent heat flux, the contrary can be observed with largest σ when the the contribution of the cold-wet patch is large, as here the largest contributions for latent heat flux can be observed.

For segment lengths of 1000 m, the relative random error for e.g. the sensible heat flux is up to 40% for the wavelet and up to 50% for the EC method. With increasing segment length, the relative error decreases. Here, one could argue that this is attributed to an improved sampling of randomly distributed updrafts and downdrafts for longer segments, resulting in less scatter. However, as σ not only depends on the number of sampled eddies for each observation but also on the sample size (the number of statistical independent flux observations), a direct comparison among σ is not permissible as the number of observed fluxes in each bin is not identical among the segment lengths shown.

Hence, to make the sampling errors comparable, we calculated the relative standard error, which is the relative standard deviation normalized by the square root of the sample size,

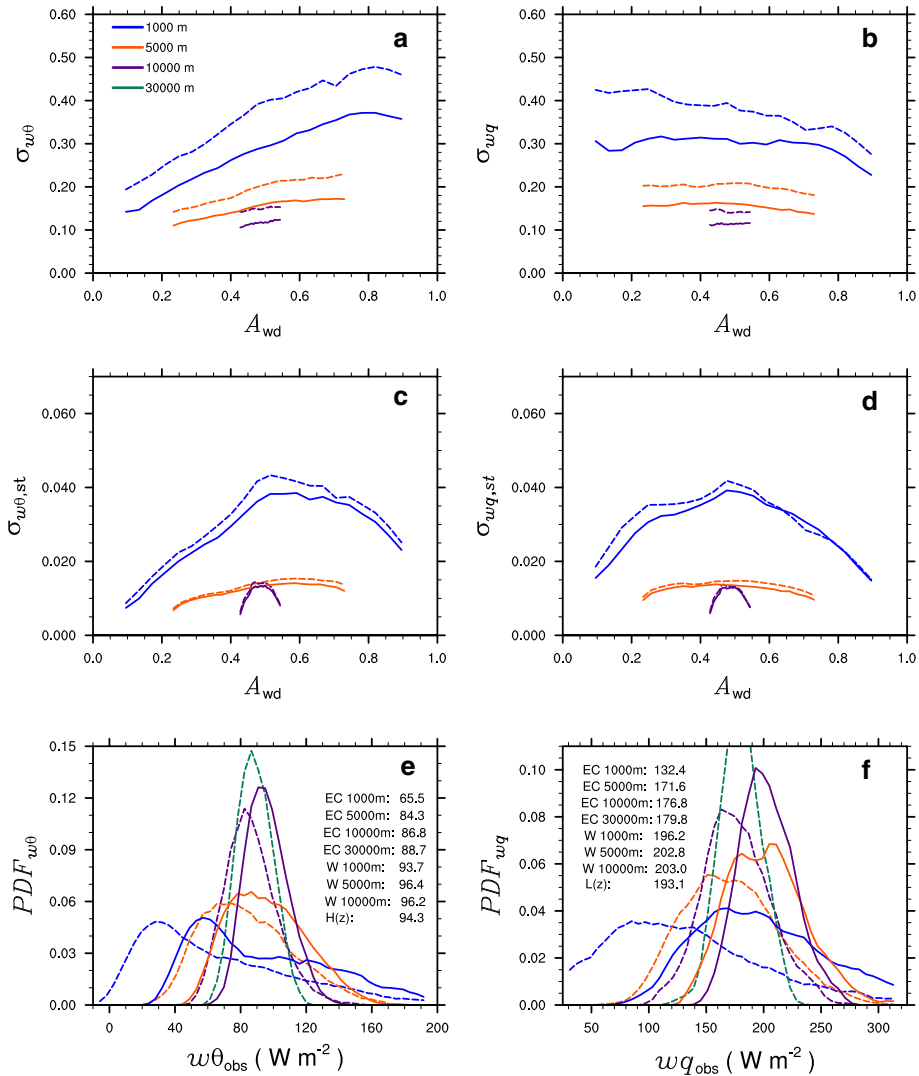


Fig. 8 **a, b** Relative standard deviation σ of observed segment fluxes according to their source-area composition for different segment lengths, **a** sensible heat flux, and **b** latent heat flux, exemplarily shown for simulation A50WL300V. Standard deviations are normalized with their respective domain-averaged surface fluxes. *Dashed* lines show EC fluxes, *solid* lines show wavelet fluxes. Please note, source-area contribution is indicated by the warm-dry patch fraction A_{wd} , the corresponding cold-patch contribution is $A_{cw} = 1 - A_{wd}$. **c, d** Relative standard error σ_{st} of the observed fluxes for **c** sensible heat, **d** latent heat. **e, f**: Probability density function (*PDF*) for **e** sensible heat, and **f** latent heat flux. In addition, the *PDF* values for 30-km segment length EC fluxes is plotted in **e** and **f**. Further, expected values for the respective *PDF*s are displayed (in text form), as well as the corresponding domain-averaged flux at the measurement level at $z = 49$ m

which in our case is the number of sampled fluxes in each bin, shown in Fig. 8c, d. First, all curves show a maximum standard error at about $A_{wd} = 0.5$ and smaller values towards larger/smaller warm-patch contributions. This can be attributed to the source-area contribution shown in Fig. 5a, with more locations showing relatively larger or smaller warm-patch

contributions and less locations with equally distributed sources, resulting in a larger number of fluxes observed at locations with larger/smaller warm-patch contributions.

If the sample size tends towards infinity, random errors vanish, as with the relative standard deviation assuming Gaussian-distributed fluxes. Hence, assuming random errors are negligible, the curves shown in Fig. 8c, d should collapse. In fact, for the longer segment lengths, the standard error is in a similar range. For the shorter segment lengths, however, the standard error shows larger values for the wavelet and EC methods in particular for measurements with similar warm- and cold-patch contributions (around $A_{wd} = 0.5$), where the sample size in each bin is significantly smaller compared to bins located at small/large A_{wd} . Among the different segment lengths, the respective standard errors for the EC and wavelet fluxes are similar, with only slightly larger values for the EC fluxes, indicating that both the EC and wavelet methods show similar random errors of the observed fluxes. Since the disaggregated wavelet fluxes (see Fig. 7) are almost independent of the applied averaging length but show similar random errors, it can thus be concluded that random errors cannot explain the large bias of the disaggregated EC fluxes for short averaging lengths.

In order to further investigate the bias of the EC fluxes, we evaluated probability density functions (PDF) for the measured segment fluxes. For short segments, one could expect a bimodal flux distribution, with maximum probability at larger/smaller fluxes for sensible heat, and smaller/larger fluxes for latent heat, over the warm-dry and cold-wet patch, respectively. For increasing segment lengths, as mentioned above, the source-area compositions tend towards equal contributions from the warm-dry and cold-wet patches, which should result in narrower PDFs with only a single maximum. Figure 8e, f shows the evaluated PDFs for different segment lengths. For the small segment length, the EC and wavelet methods show widely spread segment fluxes, which is in accordance with the larger random errors. Moreover, EC and wavelet segment fluxes are strongly skewed and show only one maximum in its distribution. As expected, with increasing segment lengths, the PDFs become narrower with a distinct maximum value and are less skewed towards smaller values. In general, the PDFs for the EC and wavelet segment fluxes exhibit a similar shape and amplitude, however, it is striking that the PDFs for the EC method are biased towards smaller fluxes for both sensible and latent heats, particularly for the small segment lengths, which becomes evident by the offset of the expected value (displayed in Fig. 8e, f in text form) from the domain-averaged flux at this height. Even for significantly larger segment lengths, here shown for 30 km, the EC-flux distributions are slightly shifted towards smaller fluxes compared to the wavelet fluxes, which indicates that the EC method exhibits some inherent bias towards smaller fluxes and is in accordance with the systematical underestimation of the disaggregated EC-based fluxes in Fig. 7. The underlying cause of this bias remains unknown, but might be related to the linear de-trending of the respective space series over each segment that acts as an additional high-pass filter (Finnigan et al. 2003). In contrast to the EC fluxes, the wavelet segment fluxes for sensible heat are much less biased and the expected values are close to the domain-averaged value, while for latent heat the wavelet segment fluxes are surprisingly slightly biased towards larger values, with increasing bias for longer segment lengths. In summary, the EC and wavelet methods show comparable random errors of the observed segment fluxes, while the EC method exhibits an increasing bias towards smaller fluxes as the segment length become shorter.

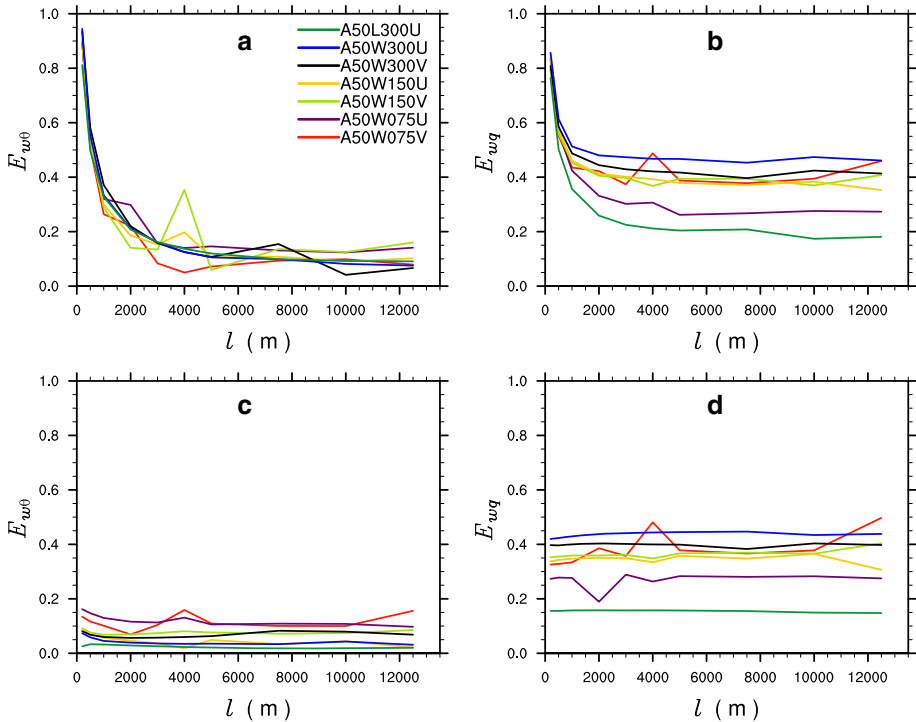


Fig. 9 Overall error in disaggregated fluxes for different segment length, **a** EC-based sensible heat flux, **b** EC-based latent heat flux, **c** wavelet-based sensible heat flux, **d** wavelet-based latent heat flux

3.6 Uncertainty of Disaggregated Patch Fluxes Depending on Segment Length

Metzger et al. (2013) considered a trade-off between the uncertainty in the measured flux due to random errors, and the ability to express a wide range of surface properties for up-scaling purposes. For their experiment they proposed a physically meaningful segment length of 1000 m. In order to investigate this trade-off in more detail, we computed the overall relative error in the disaggregated fluxes using

$$E_{w\phi} = \frac{1}{N_{\text{het}}} \sum_i^{N_{\text{het}}} \frac{|w\phi_{i,0,\text{pres}} - w\phi_{i,0,\text{da}}|}{|w\phi_{i,0,\text{pres}}|}, \quad (8)$$

with $\phi \in (\theta, q)$, $w\phi_{i,0,\text{pres}}$ and $w\phi_{i,0,\text{da}}$ being the prescribed ‘true’ and disaggregated patch fluxes, respectively, and $N_{\text{het}} = 2$ is the number of different surface types in our set-up. At this point it is important to note that while $E_{w\phi}$ is a function of the disaggregation error, it also encompasses any meteorological biases. These specifically include spatial differences in vertical flux divergence between the surface and the measurement height, and high-frequency spectral loss, which are not considered here.

Figure 9 shows the error for the disaggregated fluxes for all applied segment lengths. According to the prior discussion, the largest relative errors for the EC method can be observed for short segments with relative errors for the shortest segments exceeding 60%. Here, for both sensible and latent heat fluxes, the error decreases with increasing segment length and reaches an almost constant level at a segment length of $\approx 3\text{--}4$ km, with values ranging between

10–20% and 20–50% for the sensible and latent heat fluxes, respectively. For the wavelet method the relative error shows almost no systematic dependence on segment length, except for the shortest segment lengths where the error is slightly increased. The error is slightly smaller compared to the EC-based fluxes and ranges between 5–15% for the sensible heat flux and 15–45% for the latent heat flux.

The disaggregation error is systematically larger for the latent heat flux compared to the sensible heat flux, which correlates with the larger scatter among the segment fluxes for latent heat. Moreover, while the error spread is relatively small for the sensible heat flux among the different set-ups, the spread is rather large for the latent heat flux. However, unfortunately we were not able to draw a clear connection between the overall uncertainty of the disaggregated fluxes and patch width, amplitude of surface forcing, or occurrence of mesoscale circulation. We note that for further increasing segment lengths the source-area composition becomes more and more homogenized, and beyond some limit disaggregation via linear regression is mathematically no longer feasible. In fact, our analysis (not shown) revealed a nearly constant error up to segment lengths of 25–30 km, and largely increasing errors beyond that point.

3.7 Trade-off Between Surface Resolution and Segment Length

The LES runs used herein represent an essentially stripe-wise binary landscape, with periodically alternating surface properties. We showed that disaggregation with highly blended surface properties experienced by the observations worked well in such a landscape. For example, in the extreme case of a 12.5-km segment length highly blended fluxes of little variation associated with surface heterogeneities are able to determine the ‘pure’ patch fluxes via disaggregation. However, stripe-like surface heterogeneity is rarely observed on the Earth’s surface. To provide sufficient spatial coverage across a mesoscale model grid cell, ‘real-world’ applications typically need to consider an order of magnitude larger number of surface states. In such applications the successful solution of disaggregation depends on sufficient isolation of flux signals from the various surface patches (Metzger et al. 2013). In the current study, this can be expressed as the range of surface-property resolution $\Delta A = |A_{1\dots p,wd}|^{\max} - |A_{1\dots p,wd}|^{\min}$ (equivalent if using cold-patch contributions as well), i.e., the difference between the maximum and minimum relative contribution of the warm–dry patch occurring within all p segment fluxes used for disaggregation via Eq. 6. High ΔA (tending towards 1) indicate fine resolution of surface properties and low ΔA (tending towards 0) indicate blended surface properties, even though it is noteworthy that ΔA is only a qualitative measure and values might depend on the surface configuration. Figure 10 elucidates the relationship among the main constraints for such ‘real-world’ applications: a permissible error E at a decent resolution of surface properties, so that a proper segment length l can be chosen for the desired analysis method (eddy covariance and wavelet) to be a trade-off between these main constraints. It again becomes obvious that the range of surface heterogeneity resolution for the investigated segment lengths is quite different among the different set-ups. For the given segment lengths, the best surface-heterogeneity resolution and less blended properties can be observed for the large patch-size set-ups, while the smaller patch-size set-ups show systematically less surface resolution and tend towards larger blending of surface properties, which is in accordance with Fig. 5. For the EC method, the largest errors can be observed for fine surface resolution, which coincides with short segment lengths. For both sensible and latent heat fluxes, the error decreases rapidly at fine surface resolution and reaches an almost constant level for coarser surface resolution. For example, for the larger patch-size simulations, e.g. simulation A50WL300U/V, the error reaches an almost constant level at

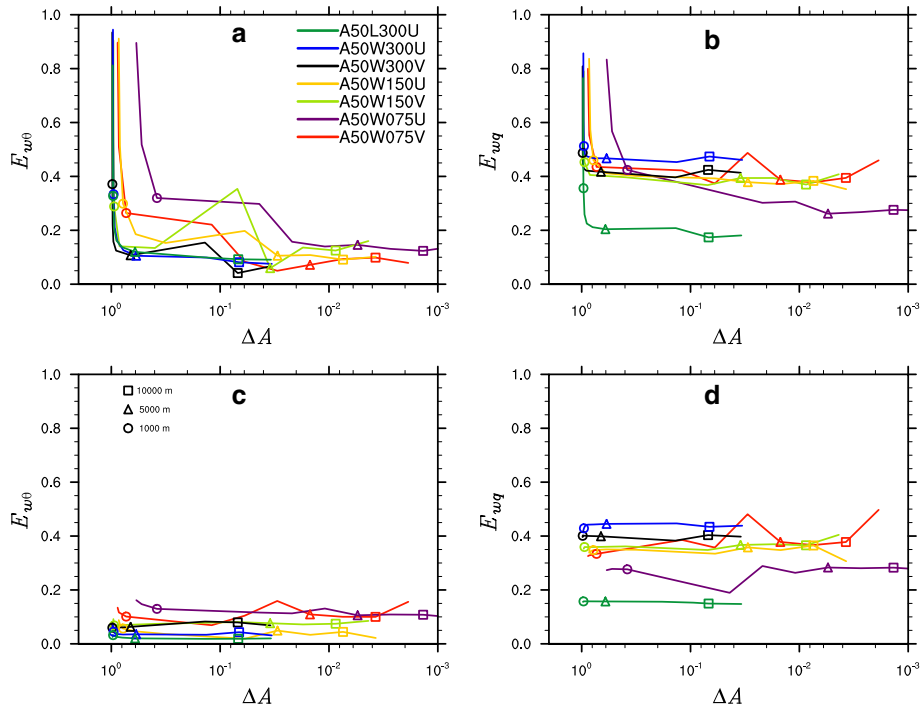


Fig. 10 Overall error in disaggregated fluxes for different surface-heterogeneity resolution ΔA , **a** EC-based sensible heat flux, **b** EC-based latent heat flux, **c** wavelet-based sensible heat flux, **d** wavelet-based latent heat flux. The coloured markers indicate the corresponding segment length at given surface-heterogeneity resolution for each set-up. For the sake of comparison with Fig. 9, please note the reversed abscissa

$\Delta A = 0.8$, where the bulk of the surface properties are still well resolved (corresponding to segment lengths of about 3–4 km). This, in turn, indicates that further increasing segment length hardly minimizes the disaggregation error at all but is only detrimental to the surface heterogeneity resolution. Hence, considering that ‘real-world’ surface heterogeneity is composed of many different scales, we recommend a segment length about 3–4 km for the EC method, which is a compromise between sufficient surface resolution and a permissible disaggregation error. For the wavelet-based fluxes, the disaggregation error is small also for fine surface resolution; thus, in terms of disaggregation error, it would be also reasonable to apply a very small segment length of 200 or 500 m.

4 Summary and Discussion

We performed a set of large-eddy simulations for idealized stripe-like surface heterogeneity with different scales and amplitudes, as well as under different flow regimes, and carried out virtual flight measurements in each of the simulations. Based on this dataset, we emulated the linear disaggregation strategy proposed by Hutjes et al. (2010) and investigated its general performance, while we compared the traditional EC method against a space–frequency (wavelet) method. For both methods, we analyzed the disaggregation error depending on the

segment length and determined the trade-off between surface blending and disaggregation error.

We showed that linear disaggregation works well in a binary landscape, invariant of patch width as well as the amplitude of surface heterogeneity, even if strong mesoscale circulations develop, confirming and extending the results of Hutjes et al. (2010). We showed, however, that the overall uncertainty of disaggregated patch fluxes depend on the method used to calculate segment fluxes, i.e., the traditional EC method or space–frequency method, with patch fluxes captured best when applying a space–frequency method. Generally, the disaggregation error was found to be smaller for the sensible heat flux than for the latent heat flux, according to less scatter of segment fluxes for sensible heat than for latent heat, which was observed, e.g., in Bange et al. (2006).

Both methods showed comparable random errors of the individual segment fluxes. The EC-based segment fluxes showed a systematic bias tending towards smaller values, particularly for small averaging lengths, while the space–frequency method showed much less biased segment fluxes. As a consequence, disaggregated patch fluxes show a general offset between the space–frequency and EC method, with smaller disaggregated patch fluxes for the EC method. For the EC method, we found that the disaggregation error strongly depends on the applied averaging length, with the largest errors for short segments where the EC method suffers most from neglecting contributions by convective updrafts and downdrafts as well as stationary mesoscale circulations. By contrast, in the case of the space–frequency method, the disaggregation error is almost invariant of the applied averaging length.

The success of ‘real world’ applications with very small-scale surface heterogeneity depends on sufficient isolation of flux signals from individual surface patches. Hence, to avoid blended fluxes and little variation of surface properties, a smaller averaging length should be preferred. In order to choose an appropriate averaging length, we quantified a trade-off between a permissible disaggregation error and surface resolution. Among our set-ups, an optimal averaging length is determined to be 3–4 km for the EC method, while with the space–frequency method an even smaller averaging length of about 200–500 m could be chosen, which nicely illustrates that flux disaggregation using the space–frequency methods is superior to the traditional EC method in terms of uncertainty as well as surface-property resolution, as already suggested by Metzger et al. (2013). In order to distinguish flux variation from different surface patches, we should use an averaging window size as small as possible, but not smaller than the time that allows emissions from the surface to reach the sensor height, and the time allowing the emission from one surface patch to fully pass the sensor. Moreover, it is worthwhile to mention that in ‘real-world’ applications footprint-model parametrizations require proper information on the turbulence statistics. Larger random errors in the turbulence statistics, due to small averaging lengths, might propagate through the footprint model into the source contribution estimates for flux disaggregation. Given these additional considerations, there might be practical merit in not reducing the segment length too much.

Besides the averaging length, another major source of uncertainty in disaggregation is the footprint. As we applied an embedded footprint model coupled with the large-eddy simulation model, information concerning heterogeneous flow conditions is already accounted for, so that the overall error in the source-area composition is minimized. Computing footprints in heterogeneous landscapes via large-eddy simulation, however, is hardly feasible in terms of computational resources, being only a tool for pure academic considerations so far. Usually, to evaluate how a signal is composed of its source and sink contributions from different surface patches, analytical models (Hsieh et al. 2000; Kormann and Meixner 2001; Kljun et al. 2002), are applied (Hutjes et al. 2010; Metzger et al. 2013; Dobosy et al. 2017), which are based on analytical solutions of the advection–diffusion equation, suffering from the lack

of general parametrization for heterogeneous flow conditions (Markkanen et al. 2010). This, in turn, raises the question concerning the uncertainty in disaggregated fluxes implied by the used footprint model, which warrants further research.

At this point we note that at the measurement level a few percent of the surface heat input is already consumed in heating and moistening the air volume between the surface and measurement level. Hence, we performed a flux-divergence correction by simply extrapolating the disaggregated fluxes to the surface by adding the difference between the domain-averaged flux at the measurement level and its domain-averaged value at the surface, which was a value constant in time. To overcome this in the ‘real world’, however, measurements should be taken at multiple heights to estimate the vertical flux divergence and correct the disaggregated fluxes accordingly. This leads to the question of how random errors (particularly for short averaging length) or time lags between the measurements propagate into the individual patch fluxes.

We applied a wavelet decomposition as a space–frequency method, noting that wavelet decomposition, however, suffers largely from edge effects when larger-scale flux contributions should be considered. To exclude edge effects, significant parts at the begin and the end of the measurement data series need to be omitted, decreasing the number of available segment fluxes used for the disaggregation. To obtain the same number of segment fluxes with wavelet decomposition as with the EC method, longer flight legs need to be flown, making analysis in the ‘real world’ more difficult considering non-stationarity due to the daily cycle. Hence, it is worth considering other space–frequency approaches, such as the empirical mode decomposition proposed by Barnhart et al. (2012), which does not require pre-determined basis functions that may result in less cut-off, reducing edge effects.

Moreover, we prescribed patch fluxes that change in a step-like manner from one patch to another. However, using a large-eddy simulation model coupled to a land-surface model, Patton et al. (2005) showed that above one surface patch the surface sensible and latent heat fluxes can significantly vary in space due to convergent and divergent flows attributed to mesoscale circulations and internal boundary layers. Furthermore, tall vegetation may add complexity to flux patterns due to flow divergence near, e.g., forest edges (Kanani-Sühling and Raasch 2015), which may be further affected by mesoscale circulations (Kenny et al. 2017). This leads to consideration of the representativeness of the patch fluxes and how disaggregation accounts for this.

We showed that the linear disaggregation method is able to determine patch fluxes emanating from surfaces with different properties. A further area of application might be of how fluxes above different surface types become blended with increasing height, i.e. the question whether a blending height exists in the convective mixed layer and how to identify it appropriately, which was already addressed in previous observational studies (Bange et al. 2006; Kang et al. 2007; Gorska et al. 2008). For example, by performing aircraft measurements at different heights and further disaggregating the measurement signals according to the underlying patches, it would be possible to calculate heat-flux profiles above different surfaces, and thus investigate the blending of fluxes.

Acknowledgements We thank the reviewers for their valuable comments and remarks that helped to improve the quality of the manuscript. All simulations were performed on the Cray XC40 at The North-German Supercomputing Alliance (HLRN), Hannover/Berlin. M. Sühling is funded by the German Federal Ministry of Education and Research (BMBF) (Grant 01LP1601A) within the framework of Research for Sustainable Development (FONA; www.fona.de). S. Metzger and D. Durden are funded by the National Ecological Observatory Network. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by Battelle Ecology, Inc. This material is based upon work supported by the National Science Foundation [Grant DBI-0752017]. Any opinions, find-

ings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. A.R. Desai and K. Xu acknowledge support of NSF AGS-1822420, Dept of Energy Ameriflux Network Management Project, Battelle Ecology, Inc. contact #3010-0401-000. NCL (The NCAR Command Language (Version 6.4.0) [software]. (2017). Boulder, Colorado: UCAR/NCAR/CISL/TDD. <https://doi.org/10.5065/D6WD3XH5>) has been used for data analysis and visualization. Simulation data and analysis scripts will be provided upon request.

References

- Bange J, Spieß T, Herold M, Beyrich F, Hennemuth B (2006) Turbulent fluxes from Helipod flights above quasi-homogeneous patches within the LITFASS area. *Boundary-Layer Meteorol* 121:127–151
- Barnhart BL, Eichinger WE, Prueger JH (2012) A new eddy-covariance method using empirical mode decomposition. *Boundary-Layer Meteorol* 145:369–382
- Beyrich F, Mengelkamp H (2006) Evaporation over a heterogeneous land surface: Eva_GRIPS and the LITFASS-2003 experiment: an overview. *Boundary-Layer Meteorol* 121:5–32
- Deardorff J (1980) Stratocumulus-capped mixed layers derived from a three-dimensional model. *Boundary-Layer Meteorol* 18:495–527
- Desjardins RL, MacPherson JJ, Neumann HH, den Hartog G, Schuepp PH (1995) Flux estimates of latent and sensible heat, carbon dioxide and ozone using an aircraft-tower combination. *Atmos Environ* 29(21):3147–3158
- Dobosy R, Sayres D, Healy C, Dumas E, Heuer M, Kochendorfer J, Baker B, Anderson J (2017) Estimating random uncertainty in airborne flux measurements over Alaskan tundra: update on the flux fragment method. *J Atmos Ocean Technol* 34(8):1807–1822
- Dolman AJ, Blyth EM (1997) Patch scale aggregation of heterogeneous land surface cover for mesoscale meteorological models. *J Hydrol* 190(3–4):252–268
- Finnigan JJ, Clement R, Malhi Y, Leuning R, Cleugh HA (2003) A re-evaluation of long-term flux measurement techniques part I: averaging and coordinate rotation. *Boundary-Layer Meteorol* 107:1–48
- Foken T, Mauder M, Liebethal C, Wimmer F, Beyrich F, Leps JP, Raasch S, DeBruin HAR, Meijninger WML, Bange J (2010) Energy balance closure for the LITFASS-2003 experiment. *Theor Appl Climatol* 101(1–2):149–160
- Gorska M, Vila-Guerau de Arellano J, LeMone M, van Heerwaarden C (2008) Mean and flux horizontal variability of virtual potential temperature, moisture, and carbon dioxide: aircraft observations and LES study. *Mon Weather Rev* 136:4435–4451
- Hsieh C, Katul G, Chi T (2000) An approximate analytical model for footprint estimation of scalar fluxes in thermally stratified atmospheric flows. *Adv Water Resour* 23:765–772
- Hutjes RWA, Vellinga US, Gioli B, Miglietta F (2010) Dis-aggregation of airborne flux measurements using footprint analysis. *Agric For Meteorol* 150:966–983
- Kanani-Sühring F, Raasch S (2015) Spatial variability of scalar concentrations and fluxes downstream of a clearing-to-forest transition: a large-eddy simulation study. *Boundary-Layer Meteorol* 155:1–27
- Kang S, Davis K, LeMone M (2007) Observations of the ABL structures over a heterogeneous land surface during IHOP_2002. *J Hydrometeorol* 8:221–244
- Kenny WT, Bohrer G, Morin TH, Vogel CS, Matheny AM, Desai AR (2017) A numerical case study of the implications of secondary circulations to the interpretation of eddy-covariance measurements over small lakes. *Boundary-Layer Meteorol* 165:311–332
- Kim J, Guo Q, Baldocchi DD, Leclerc MY, Xu L, Schmid HP (2006) Upscaling fluxes from tower to landscape: overlaying flux footprints on high-resolution (IKONOS) images of vegetation cover. *Agric For Meteorol* 146:132–146
- Kirby S, Dobosy R, Williamson D, Dumas E (2008) An aircraft-based data analysis method for discerning individual fluxes in a heterogeneous agricultural landscape. *Agric For Meteorol* 148:481–489
- Kljun N, Rotach MW, Schmid HP (2002) A three-dimensional backward Lagrangian footprint model for a wide range of boundary-layer stratifications. *Boundary-Layer Meteorol* 103:205–226
- Kormann R, Meixner FX (2001) An analytical footprint model for non-neutral stratification. *Boundary-Layer Meteorol* 99:207–224
- Kurbanmuradov O, Rannik Ü, Sabelfeld KK, Vesala T (2001) Evaluation of mean concentration and fluxes in turbulent flows by Lagrangian stochastic models. *Math Comput Simul* 54:459–476
- Lenschow DH, Mann J, Kristensen L (1994) How long is long enough when measuring fluxes and other turbulence statistics. *J Atmos Ocean Technol* 11:661–673
- Mahrt L (1998) Flux sampling errors for aircraft and towers. *J Atmos Ocean Technol* 15:416–429

- Mahrt L (2000) Surface heterogeneity and vertical structure of the boundary layer. *Boundary-Layer Meteorol* 96:33–62
- Markkanen T, Steinfeld G, Kljun N, Raasch S, Foken T (2009) Comparison of conventional Lagrangian stochastic footprint models against LES driven footprint estimates. *Atmos Chem Phys* 9:5575–5586
- Markkanen T, Steinfeld G, Kljun N, Raasch S, Foken T (2010) A numerical case study on footprint model performance under inhomogeneous flow conditions. *Meteorol Z* 19:539–547
- Maronga M, Gryschka M, Heinze R, Hoffmann F, Kanani-Sühring F, Keck M, Ketelsen K, Letzel MO, Sührling M, Raasch S (2015) The parallelized large-eddy simulation model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives. *Geosci Model Dev* 8:2515–2551. <https://doi.org/10.5194/gmd-8-2515-2015>
- Mauder M, Desjardins RL, MacPherson JI (2007) Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem. *J Geophys Res Atmos* 112:D13112. <https://doi.org/10.1029/2006JD008133>
- Metzger S, Junkermann W, Mauder M, Butterbach-Bahl K, Widemann BTY, Neidl F, Schafer K, Wieneke S, Zheng XH, Schmid HP, Foken T (2013) Spatially explicit regionalization of airborne flux measurements using environmental response functions. *Biogeosciences* 10:2193–2217
- Michaud JD, Shuttleworth WJ (1997) Executive summary of the Tucson Aggregation Workshop. *J Hydrol* 190(3–4):176–181
- Ogunjimiyo SO, Kaharabata SK, Schuepp PH, MacPherson JI, Desjardins RL, Roberts DA (1998) Methods of estimating CO₂, latent heat and sensible heat fluxes from estimates of land cover fractions in the flux footprint. *Agric For Meteorol* 117:125–144
- Oncley SP, Businger JA, Itsweire EC, Friehe CA, LaRue JC, Chang SS (1990) Surface layer profiles and turbulence measurements over uniform land under near-neutral conditions. In: 9th Symposium on boundary layer and turbulence. American Meteorological Society, Roskilde, Denmark, pp 237–240
- Ouwensloot H, Vilà-Guerau de Arellano J, van Heerwaarden C, Ganzeveld L, Krol M, Lelieveld J (2011) On the segregation of chemical species in a clear boundary layer over heterogeneous land surfaces. *Atmos Chem Phys* 11:10681–10704
- Patton E, Sullivan P, Moeng C (2005) The influence of idealized heterogeneity on wet and dry planetary boundary layers coupled to the land surface. *J Atmos Sci* 62:2078–2097
- Raasch S, Harbusch G (2001) An analysis of secondary circulations and their effects caused by small-scale surface inhomogeneities using large-eddy simulation. *Boundary-Layer Meteorol* 101:31–59
- Raasch S, Schröter M (2001) PALM—a large-eddy simulation model performing on massively parallel computers. *Meteorol Z* 10:363–372
- Samuelsson P, Tjernström M (1998) Airborne flux measurements in NOPEX: comparison with footprint estimated surface heat fluxes. *Agric For Meteorol* 98–99:205–225
- Schröter M, Bange J, Raasch S (2000) Simulated airborne flux measurements in a LES generated convective boundary layer. *Boundary-Layer Meteorol* 95:437–456
- Steinfeld G, Raasch S, Markkanen T (2008) Footprints in homogeneously and heterogeneously driven boundary layers derived from a Lagrangian stochastic particle model embedded into large-eddy simulation. *Boundary-Layer Meteorol* 129:225–248
- Strunin MA, Hiyama T (2005) Spectral structure of small-scale turbulent and mesoscale fluxes in the atmospheric boundary layer over a thermally inhomogeneous land surface. *Boundary-Layer Meteorol* 117:479–510
- Strunin M, Hiyama T, Asanuma J, Ohata T (2004) Aircraft observations of the development of thermal internal boundary layers and scaling of the convective boundary layer over non-homogeneous land surfaces. *Boundary-Layer Meteorol* 111:491–522
- Sührling M, Raasch S (2013) Heterogeneity-induced heat-flux patterns in the convective boundary layer: can they be detected from observations and is there a blending height?—A large-eddy simulation study for the LITFASS-2003 experiment. *Boundary-Layer Meteorol* 148:309–331
- Sührling M, Maronga B, Herberich F, Raasch S (2014) On the effect of surface heat-flux heterogeneities on the mixed-layer top entrainment. *Boundary-Layer Meteorol* 151:531–556
- Thomas C, Foken T (2005) Detection of long-term coherent exchange over spruce forest using wavelet analysis. *J Appl Meteorol Clim* 44:91–104
- Thomson DJ (1987) Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *J Fluid Mech* 180:529–556
- Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79:61–78
- van Heerwaarden C, Vilà-Guerau de Arellano J (2008) Relative humidity as an indicator for cloud formation over heterogeneous land surfaces. *J Atmos Sci* 65:3263–3277

- Vellinga OS, Gioli B, Elbers JA, Holtslag AAM, Kabat P, Hutjes RWA (2008) Regional carbon dioxide and energy fluxes from airborne observations using flight-path segmentation based on landscape characteristics. *Biogeosciences* 7:1307–1321
- Weil JC, Sullivan PP, Moeng CH (2004) The use of large-eddy simulations in Lagrangian particle dispersion models. *J Atmos Sci* 61:2877–2887
- Weil JC, Sullivan PP, Patton EG, Moeng CH (2012) Statistical variability of dispersion in the convective boundary layer: ensembles of simulations and observations. *Boundary-Layer Meteorol* 145:185–210
- Wicker L, Skamarock W (2002) Time-splitting methods for elastic models using forward time schemes. *Mon Weather Rev* 130:2088–2097
- Williamson JH (1980) Low-storage Runge–Kutta schemes. *J Comput Phys* 35:48–56