

“Spatial resolution and regionalization of airborne flux measurements using environmental response functions”

by S. Metzger et al.

We thank Ankur Desai for his valuable feedback on this manuscript. In below text we outline how we addressed the comments and edited the manuscript. The comments by the reviewer are indicated with an asterix (*) and are cited in italics, followed by our reply. We use **bold face font** to highlight the location of corresponding changes in the revised manuscript.

General comments

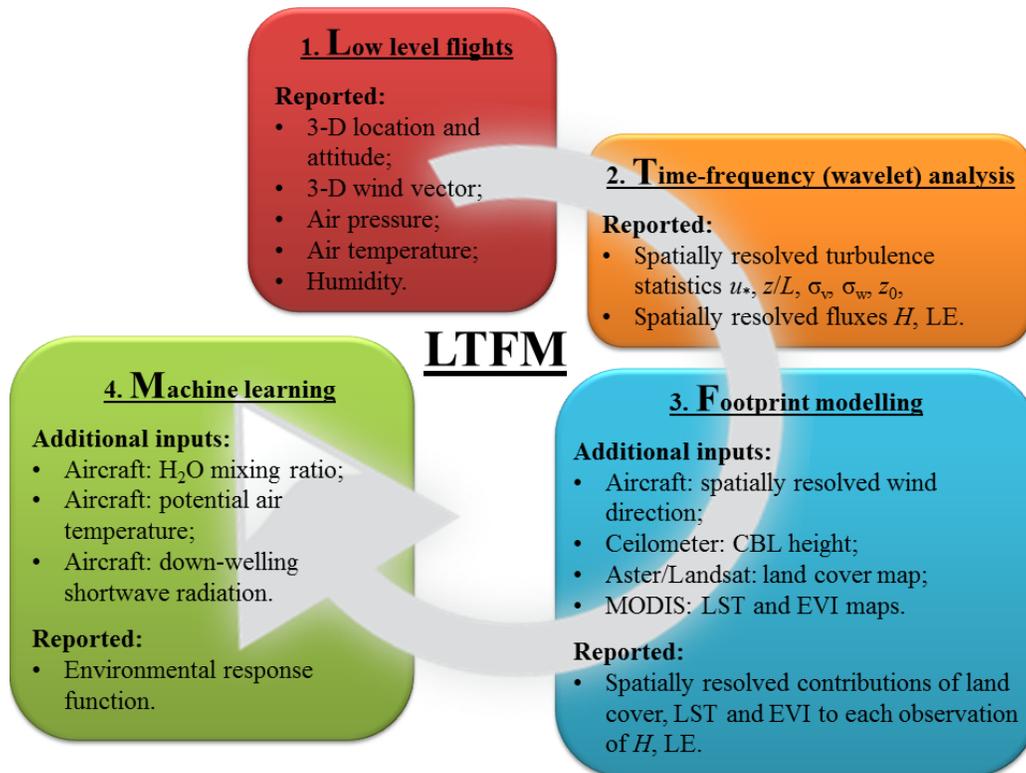
** This manuscript describes an innovative method to acquire high spatial resolution airborne eddy covariance energy fluxes in heterogeneous landscapes using a variety of techniques. For the most part, it is a methods paper, but a very important one that will be highly cited in the future. The manuscript introduces several concepts for analysis of airborne flux data including application of wavelet cross-scalogram/flux un-mixing to acquire high-resolution fluxes, application of a footprint model to these data to ascertain land surface contributions, and a boosted regression tree approach to determine catchment scale environmental response functions. A thorough uncertainty analysis for systematic and random errors is conducted. Overall, the results show some expected patterns for H and LE, but demonstrate that the techniques can produce reliable spatially explicit fluxes. The manuscript itself is lengthy as it covers a lot of ground. I have some minor comments that might help improve the flow.*

Thank you.

Specific comments

** **Comment 1-1)** It might benefit from a flowchart with a graphical description of the steps, showing the various data streams going into each step and their sources.*

On **page 59 Fig. 5** of the revised manuscript we added a flow chart. In the text of the revised manuscript on **page 5 line 13, page 16 lines 10 f., and page 18 line 20** we also added cross-references to the flow chart.



* **Comment 1-2)** *At some level, I'm still perplexed why the wavelet cross-scalogram works, but I see now that it comes at expense of high random error depending on how one defines the segment length and the importance of low-frequency contributions. Certainly, low altitude flights in the surface layer are essential for this method, as shown in the analysis of blending height and length scales. The variability in H and LE are very large across land cover types. It would have been nice to have some even literature based estimate of expected values for variation in H and LE for typical land cover types observed in this region.*

We added on **page 29 lines 3 ff. of the revised manuscript**: “These results fall well within the range of summertime ensemble average fluxes during solar noon observed by ground-based EC measurements over different land covers in this region ($100 \text{ W m}^{-2} < H < 310 \text{ W m}^{-2}$ and $100 \text{ W m}^{-2} < LE < 480 \text{ W m}^{-2}$, Gao et al., 2009; Hao et al., 2007; Hao et al., 2008; Shao et al., 2008).”

* **Comment 1-3)** *Also how about a top-down constraint? Couldn't net radiation be in principle remotely sensed? Does the estimate $H+LE$ follow patterns in R_{net} (yes I know variations in G matter)?*

It would have been useful to measure R_{net} , and also to include all radiation components in the machine learning procedure. However, in 2009 the aircraft was not yet equipped with a pyrgeometer to measure the long wave part of the down-welling radiation. In addition, to calculate the full surface energy balance (SEB), we need an estimate of the soil heat flux. In an early version of this manuscript we approximated SEB in the following way:

The downwelling and upwelling components of the shortwave radiation, $S\downarrow$ and $S\uparrow$, respectively, are measured directly by the WSMA. However the LI-2000 SZ actinometer measures solar radiation only in the visible range (400–700 nm). The measurement is internally adjusted to the spectrum of sky radiation to yield the broadband solar irradiance of a pyranometer (300–3000 nm). In consequence clouds or land covers with spectral properties different from the blue sky result in a biased measurement. Hence we compare $S\downarrow$ and $S\uparrow$ between all WSMA flights and simultaneous pyranometer measurements at three surface reference stations (Ketzer et al., 2008) On average the WSMA overestimated the radiation components by $\Delta S\downarrow=54 \text{ W m}^{-2}$ and $\Delta S\uparrow=28 \text{ W m}^{-2}$, which was corrected. The upwelling longwave radiation $L\uparrow$ was calculated from the WSMA surface temperature measurement T_s , using the Stefan Boltzmann law,

$$L\uparrow = \varepsilon \sigma_{sb} T_s^4, \quad (1)$$

with the Stefan Boltzmann constant $\sigma_{sb}=5.67 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. An emissivity $\varepsilon=1.00$ was chosen which minimizes the bias between simultaneous measurements of the WSMA and a surface reference station with comparable surface properties (Ketzer et al., 2008, heavily grazed site, KL08hg in the following). In absence of a WSMA measurement of the downwelling longwave radiation $L\downarrow$, we establish the relationship $L\downarrow=373-0.11 \cdot (S\downarrow-S\uparrow-L\uparrow)$, $R^2=0.71$ from simultaneous measurements at KL08hg. From here the net radiation Q_s^* is calculated from WSMA measurements and above parameterizations:

$$Q_s^* = S\downarrow + L\downarrow - S\uparrow - L\uparrow. \quad (2)$$

The ground heat flux GHF is modeled as a percentage of the net radiation (Liebethal and Foken, 2007). For this purpose we establish the relationship $GHF=0.17 \cdot Q_s^*$, $R^2=0.96$ from simultaneous measurements at KL08hg. Lastly the closure of the surface energy balance is calculated,

$$SEB = \frac{H + LE}{Q_s^* - GHF} \cdot 100, \quad (3)$$

with WSMA measured H , LE , in units of energy. The variability of $L\downarrow$ as well as the absolute magnitude of GHF is low. Even a large error of 15 W m^{-2} in the parameterization of $L\downarrow$ will be partially offset by the calculation of the counteracting GHF from Q_s^* . The resulting effect on the SEB is <3%.

Using this approximation and neglecting potential effects of vertical flux divergence, the surface energy balance for all flights is closed to $92 \pm 16\%$ on average. In the present version of the manuscript we did not include this assessment, because the assumptions are comparatively weak, and we thought it beyond the focus of this study.

* **Comment 1-4)** *The paper claims that ERF from un-mixing is more reliable than a model data assimilation approach, and that is possible. But there are some perplexing results. For example, it appears that solar incoming is only weakly related to H, but over the diurnal*

course, this is not the case. So is this because all the flights were mid-day only? The whole BRT and ERF approach, being somewhat new, may require a bit more discussion on how it works. In contrast, for example, the wavelet discussion is very well presented.

It was not our intention to claim that the presented methodology is more reliable than a data assimilation approach. For clarification we added at **page 4 lines 3 ff. of the revised manuscript**: “However, ERFs cannot provide insights into, e.g., ecosystem pools. Consequently ERFs might be suitable for complementing data assimilation and remote sensing approaches, e.g. through contributing to the design, constraint and evaluation of flux algorithms.”

We expanded the discussion at **page 27 lines 8 ff. of the revised manuscript**: “It appears surprising that H and LE are only weakly related to $S\downarrow$. This can be explained by using only noontime flights in the present study, where $S\downarrow$ mainly fulfils the purpose of accounting for varying cloud/radiation conditions between different measurement days. In addition, during individual flights $S\downarrow$ was usually constant to within $\leq 10\%$ (Table 5). When however using ERFs to reproduce a diurnal cycle, a much larger dependence of H and LE on $S\downarrow$ would be expected.”

We have added a flow chart on **page 59 Fig. 5 of the revised manuscript**. We hope that this will help to clarify the process of inferring an ERF from various observational data streams.

** **Comment 2-1)** The author mentions the importance of a constant AGL altitude, though given density fluctuations, wouldn't a constant pressure level be a better approach (would make the correction from T to θ less important)? Or are the WPL dry-air density corrections also being applied?*

We added at **page 8 lines 26 f. of the revised manuscript** “(ii) Measuring at a constant pressure level would make a conversion of the measured temperature and densities into potential quantities (Sect. 2.4) less important. However, over tilted or undulating terrain the aircraft would partially travel along the vertical flux gradients, thus spuriously contaminating the measured flux signal. A correction for the vertical flux gradients over complex terrain is not as straightforward and well-conditioned as the conversion into potential quantities. In order for the fluxes to remain interpretable the aircraft should thus measure at approximately constant height above terrain, i.e. the flight paths should follow the terrain contours.”, and at **page 10 lines 11 f. of the revised manuscript** “(iii) the WSMA temperature and densities are transformed to potential quantities at the mean flight altitude (pressure level) of each flight line;”.

Yes, the WPL correction is applied (page 15947 lines 27 f. of the manuscript).

** **Comment 2-2)** I'm curious why multiple heights were not flown in some cases to assess vertical flux gradients. On the other hand, I have little doubt that 50 m altitude flights should have little to no divergence.*

On four days flux measurements were performed at three altitudes and the linear vertical flux gradients for H (-0.42 – $-0.11 \text{ W m}^{-2} \text{ m}^{-1}$) and LE (0 – $0.88 \text{ W m}^{-2} \text{ m}^{-1}$) were calculated from a maximum likelihood functional relationship. In comparison, the linear vertical flux gradients

of H throughout the CBL (-0.21 – $-0.06 \text{ W m}^{-2} \text{ m}^{-1}$, Sect. 3.1 in the manuscript) are less variable and $\approx 50\%$ lower in magnitude. In all cases measured $L_R > L_H$ (Table 2 in the manuscript), thus we conclude that internal boundary layers cannot be the primary reason for this disagreement (Strunin et al., 2004). However, measurements at different heights are representative of different upwind fetches. In a heterogeneous landscape, these fetches can exhibit different source/sink strengths and mechanical properties (see response to Comment 2-4). Hence, compared to the boundary layer budget, the measured vertical flux gradients are less well constrained. Consequently we continued with the results for the boundary layer budget, and did not include above mentioned findings in the manuscript for brevity reasons. We now provide additional information at **page 18 lines 10 ff. of the revised manuscript**: “Remarkably, atmospheric pressure, z , and z_i were no significant predictors for the observed fluxes. This indicates that the chosen flight/analysis strategy effectively minimizes cross-contamination of the flux observations by vertical flux/pressure gradients.”

** Comment 2-3) Page 15942 "We hypothesize that airborne EC flux is a promising tool. . ." This is not a hypothesis. It's a claim. There are better hypotheses presented later that could be brought up here.*

We reworded the sentence at **page 5 lines 4 f. of the revised manuscript**: “We postulate that airborne EC flux measurement is a promising tool to gain new insights in the spatial variability of heat and moisture exchange across the XRC.”

** Comment 2-4) The authors a priori mask out slopes to avoid terrain generated mesoscale eddies. I'm curious on the justification and whether the data themselves indicate mesoscale eddy contribution near slopes?*

We mask out slopes to avoid cross-contaminations through (i) radiation budget effects and (ii) the mechanical generation of turbulence (Raupach and Finnigan, 1997);

- (i) The sun-exposed slope receives solar radiation more efficiently, desiccates and exhibits higher Bowen ratio over time as compared to the sun-averted slope.
- (ii) In the wake region of a hill as shallow as 10° , mechanical turbulence is increased, and an elevated shear layer, flow reversal and mean flow separation are possible (Finnigan, 1988).

At **page 8 lines 21 f. and page 9 lines 7 f. of the revised manuscript** we substituted “...orographic-induced local circulations...” with “...orographic-induced effects on radiative transfer and turbulence generation...”. Also, we now provide additional information at **page 32 lines 2 ff. of the revised manuscript**: “Masking out slopes during flight planning effectively minimizes cross-contamination of the flux observations by slope-induced effects on radiative transfer or turbulence generation. This reduces the degrees of freedom in explaining the observed flux responses, albeit potentially at the expense of oversimplifying surface-air exchange processes.”, and at **page 18 lines 13 ff. of the revised manuscript**: “Analogously the algorithm dropped elevation, aspect, and slope of the footprint modeled

source area as predictors. This shows that slope-induced effects on radiative transfer or turbulence generation do not significantly impact the flux observations.”

* **Comment 2-5)** *The KL04 cross wind footprint functions are not published yet and thus difficult to evaluate. Should include in supplement?*

We added at **page 15 lines 17 f. of the revised manuscript**: “Metzger et al. (2012) evaluated KL04+ against a backward Lagrangian reference footprint model, and good agreement was found for all considered cases.”

* **Comment 2-6)** *Page 15955 "The final BRT model. . ." re these five variables what the BRT model selects of all the variables or did the authors force the model to fit to these five only. This is not clear to me.*

We added at **page 17 lines 29 ff. of the revised manuscript**: “We use the variable dropping algorithm by Elith et al. (2008) to reach a compromise between predictive deviance and model parsimony. This algorithm (i) fits a BRT model, (ii) performs a 10-fold CV, (iii) drops the least important predictor (improvement to the model, number of splits, Friedman, 2001), and (iv) repeats this sequence until a stopping criterion is reached. The mean CV deviance can be used to decide how many variables can be removed without significantly affecting predictive performance. Here, we set an upper threshold of 30 W m^{-2} for the mean CV deviance, which equals $\leq 1/2$ the random sampling error in the flux observations (Table 4). The dropping of variables first stopped for LE at 29.2 W m^{-2} mean CV deviance, yielding a set of the five most important predictors (LST, EVI, S_{\downarrow} , MR and θ). For H the same predictor set yields a mean CV deviance of only 22.6 W m^{-2} .”

* **Comment 2-7)** *I applaud the authors for taking a serious stab at uncertainty along with the length scale analysis.*

Thank you.

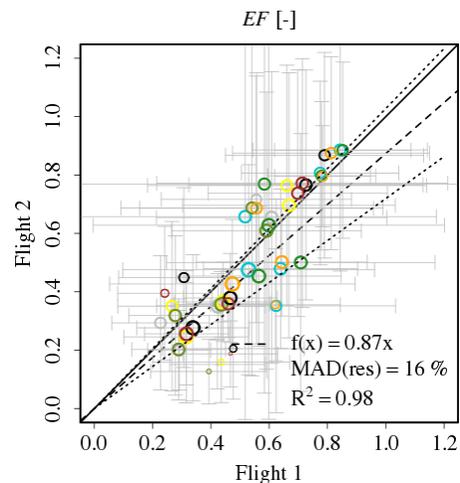
* **Comment 2-8)** *The use of continuous variables (LST/EVI) instead of land cover class is reasonable, but it assumes that all vegetation energy fluxes here can be uniquely described by these along with flight leg subset average meteorological variables. I suppose this might work here. How about somewhere where the variation might be corn crop to forest? Would it still work? The authors are very enamoured of the method, but it is important to discuss where these methods might not work or what maximum reliable altitude is possible.*

We added at **page 28 lines 4 ff. of the revised manuscript**: “Our choice of land surface and meteorological drivers appears to work well for describing the noontime surface-atmosphere exchange of heat and water vapour over a moisture-limited landscape. However, it is important to note that appropriately describing exchange processes over longer periods of time, for different landscapes or scalars might require finding an entirely different set of predictors.”

We added at **page 22 lines 27 ff. of the revised manuscript**: “The interpretation of the flux observations might be more complicated for measurement heights below the thermal blending height (limited spatial representativeness) or above the surface layer (vertical flux gradient).

** Comment 2-9) Bowen ratio is useful to look at but it has known problems when LE is small or one term is negative. Would the results change appreciably if focused on evaporative fraction ($Le/(H+LE)$)?*

We recalculated Fig. 11 (left) in the manuscript using the evaporative fraction (EF) instead of the Bowen ratio.



The median land cover specific EF agrees well between subsequent flight patterns on all measurement days. Analogously to the Bowen ratio, the land surface appears to desiccate in the course of the day. That is, during the afternoon flights 13% lower EF (vs. 12% higher Bowen ratio in the manuscript) is observed compared to the morning flights. As for the Bowen ratio, the 99.9% confidence interval includes unity slope. We conclude that the results would not change considerably when using EF.

References

Finnigan, J. J.: Air flow over complex terrain, in: Flow and Transport in the Natural Environment: Advances and Applications, edited by: Steffen, W. L., and Denmead, O. T., Springer-Verlag, Heidelberg, 183-229, 1988.

Gao, Z., Lenschow, D. H., He, Z., and Zhou, M.: Seasonal and diurnal variations in moisture, heat and CO₂ fluxes over a typical steppe prairie in Inner Mongolia, China, Hydrol. Earth Syst. Sci., 13, 987-998, doi:10.5194/hess-13-987-2009, 2009.

Hao, Y., Wang, Y., Huang, X., Cui, X., Zhou, X., Wang, S., Niu, H., and Jiang, G.: Seasonal and interannual variation in water vapor and energy exchange over a typical steppe in Inner Mongolia, China, Agric. For. Meteorol., 146, 57-69, doi:10.1016/j.agrformet.2007.05.005, 2007.

- Hao, Y., Wang, Y., Mei, X., Huang, X., Cui, X., Zhou, X., and Niu, H.: CO₂, H₂O and energy exchange of an Inner Mongolia steppe ecosystem during a dry and wet year, *Acta Oecologica*, 33, 133-143, doi:10.1016/j.actao.2007.07.002, 2008.
- Ketzer, B., Liu, H., and Bernhofer, C.: Surface characteristics of grasslands in Inner Mongolia as detected by micrometeorological measurements, *Int. J. Biometeorol.*, 52, 563-574, doi:10.1007/s00484-008-0148-5, 2008.
- Liebenthal, C., and Foken, T.: Evaluation of six parameterization approaches for the ground heat flux, *Theor. Appl. Climatol.*, 88, 43-56, doi:10.1007/s00704-005-0234-0, 2007.
- Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, *Atmos. Meas. Tech.*, 5, 1699-1717, doi:10.5194/amt-5-1699-2012, 2012.
- Raupach, M. R., and Finnigan, J. J.: The influence of topography on meteorological variables and surface-atmosphere interactions, *J. Hydrol.*, 190, 182-213, doi:10.1016/s0022-1694(96)03127-7, 1997.
- Shao, C., Chen, J., Li, L., Xu, W., Chen, S., Gwen, T., Xu, J., and Zhang, W.: Spatial variability in soil heat flux at three Inner Mongolia steppe ecosystems, *Agric. For. Meteorol.*, 148, 1433-1443, doi:10.1016/j.agrformet.2008.04.008, 2008.
- Strunin, M. A., Hiyama, T., Asanuma, J., and Ohata, T.: 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, doi:10.1023/B:BOUN.0000016542.72958.e9, 2004.

Spatially explicit regionalization of airborne flux measurements using environmental response functions

S. Metzger^{1,2,*,}, W. Junkermann¹, M. Mauder¹, K. Butterbach-Bahl¹,
B. Trancón y Widemann^{3,6,***}, F. Neidl¹, K. Schäfer¹, S. Wieneke⁴, X. H. Zheng²,
H. P. Schmid¹, and T. Foken^{5,6}**

¹Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Atmospheric Environmental Research Division, Garmisch-Partenkirchen, Germany

²Chinese Academy of Sciences, Institute of Atmospheric Physics, State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Beijing, China

³University of Bayreuth, Chair of Ecological Modelling, Bayreuth, Germany

⁴University of Cologne, Institute of Geophysics and Meteorology, Cologne, Germany

⁵University of Bayreuth, Department of Micrometeorology, Bayreuth, Germany

⁶Member of Bayreuth Centre of Ecology and Ecosystem Research, Bayreuth, Germany

* now at: National Ecological Observatory Network, Fundamental Instrument Unit, Boulder, USA

** now at: University of Colorado, Institute of Arctic and Alpine Research, Boulder, USA

*** now at: Technical University of Ilmenau, Programming Languages and Compilers, Ilmenau, Germany

Correspondence to: S. Metzger (smetzger@neoninc.org)

Abstract

The goal of this study is to characterize the sensible (H) and latent (LE) heat exchange for different land covers in the heterogeneous steppe landscape of the Xilin River Catchment, Inner Mongolia, China. Eddy-covariance flux measurements at 50–100 m above ground were conducted in July 2009 using a weight-shift microlight aircraft. Wavelet decomposition of the turbulence data enables a spatial discretization of 90 m of the flux measurements. For a total of 8446 flux observations during 12 flights, MODIS land surface temperature (LST) and enhanced vegetation index (EVI) in each flux footprint are determined. Boosted regression trees are then used to infer an environmental response function (ERF) between all flux observations (H , LE) and biophysical- (LST, EVI) and meteorological drivers. Numerical tests show that ERF predictions covering the entire Xilin River Catchment ($\approx 3670 \text{ km}^2$) are accurate to $\leq 18\%$ (1σ). The predictions are then summarized for each land cover type, providing individual estimates of source strength ($36 \text{ W m}^{-2} < H < 364 \text{ W m}^{-2}$, $46 \text{ W m}^{-2} < \text{LE} < 425 \text{ W m}^{-2}$) and spatial variability ($11 \text{ W m}^{-2} < \sigma_H < 169 \text{ W m}^{-2}$, $14 \text{ W m}^{-2} < \sigma_{\text{LE}} < 152 \text{ W m}^{-2}$) to a precision of $\leq 5\%$. Lastly, ERF predictions of land cover specific Bowen ratios are compared between subsequent flights at different locations in the Xilin River Catchment. Agreement of the land cover specific Bowen ratios to within $12 \pm 9\%$ emphasizes the robustness of the presented approach. This study indicates the potential of ERFs for (i) extending airborne flux measurements to the catchment scale, (ii) assessing the spatial representativeness of long-term tower flux measurements, and (iii) designing, constraining and evaluating flux algorithms for remote sensing and numerical modelling applications.

1 Introduction

Measurements of the exchange of heat and moisture between the land surface and the atmosphere are critical to our understanding of the role of terrestrial ecosystems in the global climate system. Ground-based eddy-covariance (EC, a summary of all notation is provided in Appendix A) measurements are suited to continuously monitor selected sites for long periods

and enable the integration in time (Baldocchi et al., 2001). However these results might only represent small areas around the immediate measurement locations (e.g. Kaharabata et al., 1997; Schuepp et al., 1992). On the other hand aircraft-based measurements can provide flux information at regional scales (e.g. Desjardins et al., 1995), but are restricted to short periods of time. Thus the temporal and spatial characteristics of ground-based and airborne measurements complement each other (Gioli et al., 2004; Mauder et al., 2007). It is desirable to integrate both approaches in an effort to provide suitable datasets for the design, constraint, and evaluation of mass and energy exchange models at site as well as at regional scales (Chen et al., 1999; Desjardins et al., 1997). In the following we briefly review the requirements for spatial scaling of airborne EC measurements, and the applicability of airborne EC measurements over complex terrain.

Aggregation approaches enable estimating the exchange over entire landscapes, provided fluxes for characteristic land cover features or domains are known (Beyrich et al., 2006). Flight path segmentation can be a useful tool to directly relate airborne EC measurements to landscape units (e.g. Desjardins et al., 1994; Vellinga et al., 2010). It is also possible to functionally relate these measurements to land cover properties, which then reflect the effects of vegetation, climate, soil and topography on the flux strength. E.g. Kirby et al. (2008) propose a method for discerning individual fluxes in a heterogeneous landscape based on subsets of “pure” flux fragments. Another approach is to utilize quantitative information about the EC measurement’s spatial context, on which basis environmental response functions (ERF, Desjardins et al., 1994) can be derived. The general idea of ERFs is to establish a relationship between spatially or temporally resolved flux observations (responses) and corresponding environmental drivers. Hence ERFs are a quantitative mechanism to extract relationships from, and to condense the information content in a dataset. If sufficiently accurate, the extracted relationships can then be used to, e.g. bridge observational scales or to adjust the spatial representativeness of ground-based flux measurements. In addition, current methods to spatially resolve surface fluxes are mainly focused on remote sensing algorithms (e.g. Fan et al., 2007) and process-based land surface models (e.g. Vetter et al., 2012). These procedures often demand far-reaching assumptions, such as the closure of the energy and water balances (e.g. Anderson et al., 2012), or are chal-

lenging with respect to the required data basis (e.g. Kaminski et al., 2012; Ziehn et al., 2011). In contrast, accurate ERFs enable inferring high-resolution surface flux maps directly from observational data with minimal, quantifiable assumptions. However, ERFs cannot provide insights into, e.g., ecosystem pools. Consequently ERFs might be suitable for complementing data assimilation and remote sensing approaches, e.g. through contributing to the design, constraint and evaluation of flux algorithms.

The forenamed applications require the relation of the airborne measured fluxes to land cover properties. To enable this requirement an aircraft is bound to measure close to the surface, where characteristic fluxes from different land covers are not yet fully homogenized (or blended, Mason, 1988; Wood and Mason, 1991). Moreover the flux must be measured at a constant altitude above ground, so as to avoid artificial flux contributions through altitude fluctuations along vertical gradients (Vickers and Mahrt, 1997). However investigation areas are seldom ideally flat, and topography can vary significantly throughout a domain. To safely follow terrain contours at a low and constant altitude above ground, the aircraft must possess a low ratio of true airspeed to climb rate. Only a few airborne platforms fulfil this requirement (e.g. Bange et al., 2006; Gioli et al., 2004; Thomas et al., 2012), with the weight-shift microlight aircraft (WSMA) being one of them (Metzger et al., 2011, 2012). In forenamed studies we describe a WSMA that enables airborne EC flux measurements in remote settings at reasonable cost and minimal infrastructural demand. The objectives of the present study are to investigate the possibilities of (i) deriving meaningful EC fluxes from WSMA measurements over complex terrain, and (ii) scaling the results to a domain of interest.

We applied the WSMA over the undulating steppe of the Xilin River Catchment (XRC), Inner Mongolia, China (Fig. 1). On 21 days in the summer of 2009 flights along line transects were conducted at 50–100 m a.g.l. From boundary layer scaling it is found that the vertical flux gradients below the flight level satisfy the surface layer definition (constant within $|5-10\%|$). Hence measured sensible- (H) and latent heat flux (LE) can be interpreted as surface fluxes (Sect. 3.1). Because of its climate and management practices typical for semiarid grasslands of China (Butterbach-Bahl et al., 2011), intensive ecological research has commenced in the XRC from the late 1970s (Jiang, 1985). Besides Tibet, Inner Mongolia is Chinas most important

province for grassland based livestock production, and desertification due to overgrazing is a major problem for extensive areas (Meurer and Jiang, 2001). In addition the land cover in the investigation area varies distinctly in space and time (Ketzer et al., 2008; Schaffrath et al., 2011). We **postulate** that airborne EC flux measurement is a promising tool to gain new insights in the spatial variability of heat and moisture exchange across the XRC.

In the present paper we firstly introduce the WSMA and the on-board measurements, and give an overview of the climate and physical composition of the study area (Sects. 2.1, 2.2). We then describe a measurement strategy (Sect. 2.3) which is linked to a novel data processing approach (Sect. 2.4). A wavelet transformation allows us to resolve fluxes above each overflow cell of a 90 m land cover raster without neglecting flux contributions on much larger scales. In combination with footprint modelling and a non-parametric machine learning technique, an ERF is computed between the airborne flux observations and meteorological and land surface drivers (**Fig. 5 provides an overview of the data flow**). In Sect. 3 we present the results of this functional relationship and evaluate its potential to explain the spatial distribution of the heat and moisture exchange along the flight lines. We interpret the results in the context of blending scales and statistical errors, and discuss the uncertainty associated with using the ERF to predict fluxes to freely selectable domains in the XRC (Sect. 3.2.4). Lastly, we give an outlook on potential applications of WSMA flux measurements and future improvements of the presented methodology (Sect. 4).

2 Materials and methods

2.1 The weight-shift microlight aircraft

The structure of a WSMA differs from common fixed-wing aircraft: it consists of two distinct parts, the wing and the trike, which hangs below the wing and contains pilot, engine and the majority of the scientific equipment. This particular structure provides the WSMA with exceptional transportability and climb rate, which qualifies it for applications in inaccessible and topographically structured terrain. A detailed description of the physical properties of the

WSMA used in this study as well as characteristics and manufacturers of sensors and data acquisition is given in Metzger et al. (2011, 2012). In short: most variables are sampled at 100 Hz and are block-averaged and stored at 10 Hz, yielding a horizontal resolution of approximately 2.5 m. In this study, we use the 10 Hz measurements of the three dimensional wind speed ($\sigma_w = 0.04 \text{ ms}^{-1}$ precision), temperature ($\sigma = 0.04 \text{ K}$), humidity ($\sigma = 0.005 \text{ gm}^{-3}$), and the height a.g.l. ($\sigma = 0.04 \text{ m}$). From error propagation it was found that changes in friction velocity (u_*), H and LE of 0.02 ms^{-1} , 5 W m^{-2} , and 3 W m^{-2} , respectively, can be reliably distinguished (Metzger et al., 2012). In addition, we use for this study slow measurements ($\leq 0.1 \text{ Hz}$) of humidity (TP3 dew point mirror, Meteolabor AG, Wetzikon, Switzerland), surface temperature (CT infrared thermometer, Optris GmbH, Berlin, Germany), and down-welling short wave radiation (LI-200 SZ, LI-COR Inc., Lincoln, Nebraska, 400–1100 nm, within an error of 5 % equal to pyranometer measurements, 300–3000 nm).

2.2 Study area

Airborne EC flux measurements were performed in the XRC from 23 June to 4 August 2009. The hilly investigation area lies south of the provincial capital Xilinhot, Inner Mongolia, China ($43.1\text{--}43.9^\circ \text{ N}$, $116.0\text{--}117.2^\circ \text{ E}$, 1000–1500 m a.s.l., Fig. 2). The XRC covers an area of $\approx 3670 \text{ km}^2$ and is characterized by temperate continental monsoon climate, with cold and dry winters and warm and wet summers. From data of the years 1982–2005 at the Inner Mongolia Grassland Ecosystem Research Station (IMGERS, 43.63° N , 116.70° E , 1187 m a.s.l., Fig. 2) the monthly mean air temperature ranges from -21° C in January to $+19^\circ \text{ C}$ in July, with an annual mean of $+1^\circ \text{ C}$ (Liu et al., 2008). Variability in total annual precipitation is high (166–507 mm) with a mean annual sum of 335 mm. Typically 60–80 % of the rainfall occurs from June to August (Chen, 1988). June 2009 (57 mm) and August 2009 (60 mm) were in the usual range, but July 2009 (35 mm) only received half of the long-term average rainfall. Detailed information on the meteorological conditions during the flight campaign is provided in Appendix B. Chestnut soils are the main zonal soil types, with a land cover dominated by *Stipa grandis*, *S. krylowii*, *Artemisia* and *Leymus chinensis* steppe. Throughout the XRC the abundance of C4 species in the steppe composition is relatively homogeneous (15–25 %, Auerswald

et al., 2009). The growing season usually lasts from the end of May to late September (Liang et al., 2001).

Land cover in the XRC had been classified on the basis of a Landsat 7 Thematic Mapper image of 17 August, 2005 (Wiesmeier et al., 2011). In recent years however the development of settlements sprawled, and irrigated agriculture is gaining popularity (Qi et al., 2007). The Bowen ratio B_o of the latter is distinctly different from the land cover classes that already exist in the classification of Wiesmeier et al. (2011). This land cover classification was thus updated and extended by visual reclassification of Advanced Spaceborne Thermal and Reflection Radiometer (ASTER) images of 7 and 28 April, 2009. The result is a land cover map with a resolution of 90 m, which is dominated by generic steppe (71 % coverage), intersected by a dune belt (10 %, Fig. 2). The coverage of bare soil, mountain meadow, marshland and rainfed agriculture is each ≈ 5 %, and the coverage of water bodies, settlements and irrigated agriculture is sub-per cent. In the context of this study the land cover classification represents the longer-term effects of vegetation, climate, soil and topography.

The spatial variation of temperature and precipitation in the XRC follows altitudinal and latitudinal trends (Auerswald et al., 2009; Wittmer et al., 2010). To resolve the effective state of biophysical surface properties over time we use Moderate Resolution Imaging Spectroradiometer (MODIS) data. We chose 8-day composites of the **daytime** land surface temperature (LST, MOD11A2.5, 1 km resolution), and 16-day composites of the enhanced vegetation index (EVI, MOD13Q1, MYD13Q1, 250 m resolution) for this purpose. Due to an 8-day overlap of the EVI data products by the MODIS Terra and Aqua missions, LST and EVI datasets could both be acquired for 4, 12, 20, 28 July, and 5 August 2009. The LST and EVI datasets were bi-linearly interpolated to the 90 m resolution of the land cover classification, and linearly interpolated in time to yield an individual map for each flight day. The spatial gradients in temperature and precipitation throughout the XRC are clearly reproduced by the greenness of the vegetation (Fig. 2). The spatio-temporal resolution of the MODIS data enables assessing the actual state of the biophysical conditions at the land surface. LST and EVI vary significantly throughout the study period, but also between the different land cover types (Fig. 3). All land cover types follow a similar temporal trend, with LST and EVI peaking mid-July and end-July, respectively. While

open water is the coolest surface, LST as well as greenness increase from mountain meadow over marshland to irrigated agriculture, which are likely strong sources of evapotranspiration. The reverse relationship (increasing LST and decreasing EVI) is found for settlements, rainfed agriculture, dunes, steppe and bare soil, which are likely strong sources of H .

5 A ceilometer (LD40 – Vaisala Oy, Helsinki, Finland) was deployed at IMGERS and provided vertical profiles of the atmospheric laser radiation backscatter intensity (Mini light detection and ranging, originally applied for the detection of the cloud base height). The depth of the convective boundary layer (CBL) was inferred from 10 min means of these data, in combination with semi-daily radiosonde ascends in nearby Xilin Hot (World Meteorological Organization station 10 54102, <http://weather.uwyo.edu/upperair/sounding.html>). For this purpose the maximum gradient method is used, which enables the detection of up to five lifted inversions (Emeis et al., 2008; Helmis et al., 2012; Munkel and Roininen, 2010). It is assumed that the aerosol number concentration, size distribution, shape and chemical composition (refractive index, absorption) adapt rapidly to the CBL structure. If there was more than one maximum or layer detected, the 15 lowest one is taken as the CBL depth. The CBL depth is used in Sect. 2.4.1 for the calculation of atmospheric length scales and in Sect. 2.4.3 for the calculation of the airborne flux measurements' source areas. In addition the cloud cover during the flight periods were monitored by ground personnel at IMGERS.

2.3 Measurement strategy

20 Advancing into more complex terrain, a flight strategy needs to be derived that considers (i) pilot safety, (ii) vertical flux gradients, (iii) orographic-induced **effects on radiative transfer and turbulence generation**, (iv) statistical errors, and (v) the land cover distribution. Such flight strategy was derived for the XRC study region using the geographic information system ArcMap 9.2 (Environmental Systems Research Institute, Redlands, USA). (i) A minimum flight level of 25 50 m above ground was found to provide the pilot with sufficient clearance for safe flight even under buoyancy driven turbulence. **(ii) Measuring at a constant pressure level would make a conversion of the measured temperature and densities into potential quantities (Sect. 2.4) less important. However, over tilted or undulating terrain the aircraft would partially travel along**

the vertical flux gradients, thus spuriously contaminating the measured flux signal. A correction for the vertical flux gradients over complex terrain is not as straightforward and well-conditioned as the conversion into potential quantities. In order for the fluxes to remain interpretable the aircraft should thus measure at approximately constant height above terrain, i.e. the flight paths should follow the terrain contours.

(iii) We use a digital elevation model (Shuttle Radar Topography Mission, Tile 60_04, data version 4.1, Jarvis et al., 2008) to calculate slope angles. In an effort to avoid immediate orographic-induced effects on radiative transfer and turbulence generation all locations within 500 m radius around slopes exceeding 6° were masked. This radius approximately equals five times the standard deviation (SD) of the terrain elevation. (iv) To reduce the statistical errors the flight path should be long and perpendicular to the mean wind direction. Thus we aligned straight flight lines along four wind axes in the areas which were not masked in the previous step. (v) Of the flight lines that were frequently perpendicular to the mean wind direction, those that best represented the land cover distribution in the XRC were covered on multiple flight days. This strategy results in a terrain-following flight patterns, with typical altitude gradients of 100 m vertical on 10 km horizontal. The climb angles of the aircraft rarely exceeds $\pm 5^\circ$, and on average the height above ground is constant to within 12 m (Table 1). Each flight line was repeated until a minimum of 40 km of data was acquired.

The aircraft was operated from IMGERS (Fig. 2). For the present study we use data from six days in July 2009 (Table 1), which were selected according to the availability of auxiliary datasets and homogeneity of the down-welling shortwave radiation S_\downarrow along the flight tracks (Table 5). On each day measurements were carried out along pairs of approximately parallel flight lines at a nominal airspeed of 27 ms^{-1} , with eight individual flight lines in total. Each pair is located across or along the humidity and temperature gradients in the XRC (Fig. 2). This strategy provides two independent datasets for each flight day, and covers the fundamental climatic gradients in this area.

The land cover type most frequently observed below all flight lines is steppe (Table 1). Flight lines with significant surface coverage of marshland or irrigation tend to be greener (higher EVI) compared to flight lines with significant coverage of dunes or bare soil. In the following, we investigate whether spatially resolved land cover information can be used as predictor for H and

LE measured along the flight lines. At this we hypothesize that LST and EVI are representative proxies for heat and moisture sources on the surface, respectively (e.g. Glenn et al., 2008; Lyons and Halldin, 2004; Nagler et al., 2007). Because aircraft measurements cover a broad state space, the resulting observations are particularly suited to infer ERFs.

5 2.4 Data processing

An analysis package for the processing of the airborne EC data was developed in GNU R version 2.13 (R Development Core Team, 2012). The analysis package is described in detail in Metzger et al. (2012) and is available upon request. Relevant processing steps are (i) the raw data is screened for spikes; (ii) humidity from fast response and slow reference sensors are merged using a complementary filter; (iii) the WSMA temperature and densities are transformed to potential quantities at the mean flight altitude (pressure level) of each flight line; (iv) the time delay due to separation between the vertical wind measurement and the temperature and humidity measurements is corrected by maximising their lagged correlation; (v) the WPL correction according to Webb et al. (1980) is used to correct LE for density fluctuations; (vi) correction for spectral artefacts ($-1 \pm 1 \%$, $-2 \pm 1 \%$, and $-6 \pm 2 \%$ for the SDs in the wind components along, transverse and vertical to the aircraft coordinate system, respectively) and cospectral artefacts ($-4 \pm 4 \%$, $0 \pm 1 \%$, and $0 \pm 2 \%$ for u_* , H and LE, respectively) after Metzger et al. (2012); (vii) calculation of random and systematic statistical errors after Lenschow and Stankov (1986), Lenschow et al. (1994).

20 2.4.1 Horizontal mixing between surface and flight level

Horizontal mixing between the surface and the flight level results in the spatial integration of fluxes above heterogeneous terrain, a process also referred to as “blending” (e.g. Mason, 1988). Working toward an ERF between surface properties (driver) and flux measurement (response), we will test three hypotheses related to horizontal mixing. At flight level (i) the turbulence is in approximate equilibrium with the land surface in the flux footprint; (ii) the measured turbulence statistics are representative of the mechanical setting upwind; (iii) changes in the turbulent flux

can be resolved at the horizontal scale of surface heterogeneities.

Several analytical formulations have been developed to characterize a mixing regime (e.g. Mahrt, 2000; Raupach and Finnigan, 1995; Wood and Mason, 1991). Such formulations are usually based on the comparison of characteristic length scales for surface heterogeneity and CBL mixing. Here we use the autocorrelation function to estimate the typical horizontal scale of surface heterogeneity L_H . For this purpose we integrate the autocorrelation function of the WSMA measured surface temperature T_s at distance d along a flight line from zero lag to the first crossing with zero at lag c_0 :

$$L_H = \int_0^{c_0} \frac{\overline{T'_s(d)T'_s(d+c)}}{\overline{T'_s(d)^2}} dr, \quad (1)$$

where overbars denote the mean along a flight line, and primes denote the deviations from this mean. L_H can be interpreted as the spatial coherence of surface features along the flight path (e.g. Strunin et al., 2004). In the following we assume that L_H is isotropic within the flux footprint, i.e. also representative perpendicular to the flight path.

In order to further characterize the mixing regime, L_H can be compared to atmospheric length scales. These length scales inter-relate the transport strengths in the horizontal and vertical directions, and correspond to the along wind distance after which the air mass below a reference level is approximately homogenized. Formulations for atmospheric length scales mainly differ in their use of (i) the measures of transport strengths, and (ii) the vertical or horizontal reference scale. Raupach and Finnigan (1995) proposed a length scale L_R (now also referred to as Raupach length), which characterizes the mixing regime throughout the entire CBL:

$$L_R = 0.8z_i \frac{\bar{u}}{w_*}, \quad (2)$$

with CBL depth z_i , average (bulk) horizontal transport velocity u , and convective velocity w_* . Because u cannot be measured directly, it is substituted in Eq. (2) and in the following investigations with the measured horizontal wind speed at flight altitude. The influence of surface heterogeneities with spatial scales L_H that are small compared to L_R is confined below the CBL

top. In this case the concept of a “blending height” within the CBL arises. The blending height corresponds to a vertical level at which the turbulent flow field over heterogeneous terrain approaches equilibrium with the local vertical gradient. If the blending height is confined within the surface layer, Monin-Obukhov similarity can be applied above the blending height (Mahrt, 2000). Wood and Mason (1991) define the thermal blending height for unstable stratification,

$$z_{\text{TB1}} = L_H \frac{\overline{w'\theta'_{0,v}}}{\overline{u}\theta_{0,v}} \cdot 323, \quad (3)$$

with the buoyancy heat flux from the virtual potential temperature $\theta_{0,v}$, and the horizontal wind speed u at the blending height. This thermal blending height can be rearranged as thermal blending length,

$$L_{\text{TB1}} = \bar{z} \frac{\overline{u}\theta_{0,v}}{w'\theta'_{0,v}} \cdot 3.1 \cdot 10^{-3}, \quad (4)$$

now representing the smallest scale of surface heterogeneity that significantly influences the turbulent flow at flight level z above ground. An improved version of the thermal blending length was proposed by Mahrt (2000),

$$L_{\text{TB2}} = L_R \frac{\overline{\theta_{0,v}}}{\sigma_{T_s}} \cdot 4.3 \cdot 10^{-3}, \quad (5)$$

which considers the SD of T_s as measure for the amplitude of surface heterogeneity. The numerical factors in Eqs. (3)–(5) were estimated from observations by Mahrt (2000). In Sect. 3.1 the results of above formulations are used to test the initial hypotheses related to horizontal mixing.

2.4.2 Wavelet cross-scalogram

- 5 The differentiation of land cover types is at odds with the classical time-domain EC method, which assumes homogeneous terrain. However, Parseval’s Theorem implies that the covariance of signals may be studied not only in the time domain, but equivalently in the frequency domain.

There, we have the wavelet transform family of methods at our disposal, which are particularly suited for the spectral analysis of non-stationary signals.

A wavelet is a signal that is localized in both time and frequency. Different localizations of the same basic shape (daughter wavelets) are constructed as a function of time t by defining

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) = \frac{1}{\sqrt{|a|}} \psi(q), \quad (6)$$

where ψ is a suitable mother wavelet, a is a scale parameter (in frequency domain), b is a location parameter (in time domain), and $q = (t-b)/a$ is a dimensionless coordinate (in time-frequency space). The convolution $\int x(t)\psi_{a,b}(t)dt$ of a signal x with a daughter wavelet $\psi_{a,b}$ yields a wavelet coefficient $W_x(a,b)$, a wavelet transform being a collection of such coefficients. We follow the procedure of Torrence and Compo (1998), using the continuous wavelet transform approximation for discrete input. The chosen mother wavelet is the Morlet wavelet $\psi(q) = \pi^{-1/4} e^{i\omega_0 q} e^{-q^2/2}$ with the frequency parameter $\omega_0 = 6$. The relevant parameters are spaced exponentially in frequency and uniformly in time, respectively: $a_j = a_0 2^{j\delta j}$ for $j = 0, \dots, J$ and $b_n = n\delta t$ for $n = 0, \dots, N-1$, with length of the data set N , initial scale parameter $a_0(\delta t, \psi)$ and number of scale increments $J(a_0, \delta j, \delta t, N, \psi)$. a_0 and J are chosen such that the extreme wavelet scales match the period of the Nyquist frequency, here 0.2 s, and the duration of the dataset, respectively. The unit of increment in the time domain, δt , is given by the sampling period of the time series, here 0.1 s. The unit of increment in frequency domain, δj , can be set to different values, with smaller values increasing both resolution and redundancy. For the present data, the results of the wavelet analyses were insensitive to the choice of δj in the range $0.0625 < \delta j < 0.25$. Hence we follow the example of Torrence and Compo (1998) and use $\delta j = 0.125$. The wavelet scalogram of a signal x is defined as the matrix of $|W_x(a,b)|^2$ for all admissible a, b . Likewise, the wavelet cross-scalogram of two signals x, y is the matrix of $W_x(a,b)W_y(a,b)^*$, where $*$ denotes the complex conjugate. The global covariance of x and y can be estimated by weighted averaging

$$\text{cov}_{a,b} = \frac{\delta j \delta t}{C_\delta N} \sum_{j=0}^J \sum_{n=0}^{N-1} \frac{W_x(a_j, b_n) W_y(a_j, b_n)^*}{a_j}, \quad (7)$$

where C_δ is a reconstruction factor specific to each mother wavelet, here 0.776 for the Morlet wavelet. The covariance can also be estimated locally for a subinterval of either j or n . This is a useful feature for dealing with changes in land cover: the continuous wavelet transform is highly redundant, with high correlation between adjacent low-frequency coefficients. Therefore, the covariance for a subinterval in time can be estimated without neglecting low-frequency, large-scale contributions. The downside of the large support of low-frequency wavelets is that edge effects due to the finite overall data set increase with scale. Torrence and Compo (1998) define the cone of influence (COI) as the boundary where the power of edge-related artefacts is damped by a factor of e^{-2} . Integration over all scales yields results close to the time-domain EC method (here, -7 to -3 % median differences), but also includes less reliable estimates above the COI (e.g. Strunin and Hiyama, 2004). Considering only scales below the COI rejects those less reliable estimates (e.g. Mauder et al., 2007). However, because part of the scale range is excluded, such procedure also systematically increases the discrepancy between Wavelet and time-domain EC methods (here, -22 to -7 % median differences). Moreover, the COI tapers toward the centre of the dataset (Fig. 4). Different scales would be included when estimating the covariance for subintervals, depending on the position along the flight path. Hence, for the localization of flux contributions in space we (i) integrate over all scales of a subinterval, and (ii) use a correction factor for each individual flight path to compensate the difference between Wavelet and time-domain EC methods. Such procedure is suited for the derivation of ERFs, because (i) it enables localization in space, (ii) it considers contributions to the local flux from scales that are larger than the subinterval, (iii) it is not biased with respect to the global time-domain covariance, and (iv) uncertainty arising from edge-effects is propagated in the ERF and included in the final uncertainty metric (Sect. 2.4.4).

Computations were performed with a continuous wavelet transform package written in GNU R (R Development Core Team, 2012), partially based on the published code of Torrence and Compo (1998) available from <http://atoc.colorado.edu/research/wavelets>. u_* , H , LE (and analogously the SDs of the wind components) are calculated for overlapping subintervals of 1000 m length. The subintervals are centred above each cell of the land cover/LST/EVI grids that was overflowed by the WSMA, principally yielding one flux observation every 90 m. The resulting

sample size for all 12 flights in Table 1 is $N = 8446$.

2.4.3 Footprint modelling

The footprint- or source weight function quantifies the spatial contributions to each measurement (Schmid, 2002; Vesala et al., 2008). For this purpose we use the footprint model of Kljun et al. (2004, KL04), which is a parameterisation of the backward Lagrangian model of Kljun et al. (2002) in the range $-200 \leq z/L \leq 1$, $u_* \geq 0.2 \text{ ms}^{-1}$, and $1 \text{ m} \leq z \leq z_i$. The parameterisation depends upon u_* , measurement height z , SD of the vertical wind σ_w and the aerodynamic roughness length z_0 , of which u_* , z and σ_w are measured directly by the WSMA. The roughness length is inferred using the logarithmic wind profile with the integrated universal function for momentum exchange after Businger et al. (1971) in the form of Höögström (1988). The KL04 is a cross-wind integrated footprint model, i.e. it does not resolve the distribution perpendicular to the main wind direction. In order to account for cross-wind dispersion the KL04 was combined with a Gaussian cross-wind distribution function (Kljun et al., 2013, in the following referred to as KL04+). In addition to above variables the SD of the crosswind from WSMA measurements and the depth of the CBL z_i from ceilometer measurements (Sect. 2.2) are used. This results in a computationally fast footprint parameterisation which considers 3-D dispersion and is not constrained to applications in the surface layer. Metzger et al. (2012) evaluated KL04+ against a backward Lagrangian reference footprint model, and good agreement was found for all considered cases.

Turbulence statistics for a 1000 m long subinterval over the wavelet scalograms (Sect. 2.4.2) are used to evaluate the KL04+. One evaluation is carried out for each overflow cell of the land cover, LST and EVI grids (i.e. every 90 m along the flight path). With the overflow grid cell as base point, the footprint weights w_{xy} ($\sum w_{xy} = 1$) are calculated for each grid cell with

position x, y , relative to the base point. From here the footprint composition is calculated,

$$\text{LST} = \sum_x \sum_y w_{xy} \cdot \text{LST}_{xy}, \quad (8)$$

$$\text{EVI} = \sum_x \sum_y w_{xy} \cdot \text{EVI}_{xy}, \quad (9)$$

with the land surface temperature and enhanced vegetation index for each grid cell, LST_{xy} and EVI_{xy} , respectively. For graphical representation all evaluations of KL04+ along a flight line are superimposed and normalized to a sum of unity. Additional information and references regarding footprint calculations along line transects can be found in Hutjes et al. (2010) and Meijninger et al. (2006).

2.4.4 Environmental response function

We base the development of a catchment-specific ERF on the works of Chen et al. (1999), Hutjes et al. (2010) and Ogunjemiyo et al. (2003). The general idea is to establish a functional relationship between spatially or temporally resolved flux observations (responses) and corresponding environmental drivers. Figure 5 provides an overview of a novel approach to ERF presented in the following.

Thus far, a suitable number of flux observations was obtained by either shortening the time-domain EC averaging interval (Chen et al., 1999; Ogunjemiyo et al., 2003), or by stratifying repeated observation along the same flight line on different days (Hutjes et al., 2010). The inherent drawbacks are the neglect of either long wavelength contributions to the flux measurement, or inter-day variability of ecosystem drivers. Both are overcome using the wavelet cross-scalogram technique (Sect. 2.4.2).

Previously, the development of ERFs has solely focused on drivers in the footprint of the flux observations, namely discrete land cover classifications. This procedure ignores within-class variability across a catchment, e.g. along climatic or altitudinal gradients, which can be overcome by using continuous variables such as LST and EVI instead. In addition, the present

approach considers the meteorological drivers $S\downarrow$, mixing ratio (MR), and potential temperature (θ). This avoids the need to stratify or pre-select data, and enables constructing a single ERF that is valid for the entire observation period and, within in range of the measured variables, throughout a catchment of interest.

5 Hitherto, ERFs were determined as the inverse of a linear mixing matrix, using either numerical (Chen et al., 1999) or regression methods (Hutjes et al., 2010; Ogunjemiyo et al., 2003). Such procedure assumes a linear relationship between drivers and responses, which is subject of on-going discussion and research (e.g. Raupach and Finnigan, 1995). Instead, the present approach uses boosted regression trees (BRT), a non-parametric machine learning technique,
10 to establish an ERF between drivers and responses. In contrast to parametric approaches, BRT does not assume a predetermined form of the response, but constructs an ERF according to information in the data. **It is for this reason that not the absolute values of the land surface and meteorological drivers are important, but rather their spatial variability and coherence. In case of the land surface drivers for example, the only assumption made here is that the spatial patterns of LST and EVI approximate the spatial patterns of source strength in H and LE (e.g., Holmes, 1970; Oke, 1987). This is a much weaker assumption than a mechanistic link, and adds power to the method.** BRT can fit complex nonlinear relationships, automatically handle interactions between drivers, and provide predictive performance that is superior to most traditional modelling methods (e.g. Hu et al., 2010). Here we use the BRT work package by Elith et al. (2008), which
20 builds upon the GBM library by Ridgeway (2012). To identify the optimal choice of parameters and variables for the BRT, a sensitivity analysis was conducted using the cross-validation (CV) procedure described in Elith et al. (2008). During cross validation all available data is divided into 10 random combinations of training (90 %) and evaluation (10 %) fractions, which allows assessing and optimising model performance. The parameter setting that minimized predictive
25 deviance for the present dataset were found to be, respectively; absolute (Laplace) error structure, bag fraction (0.7), tree complexity (5), learning rate (0.1), and number of trees (10^4). The initial set of variables also included time of the day, MODIS albedo, atmospheric pressure, land cover, z , z_i , u , u_* , z_0 , virtual potential temperature, as well as elevation, topographic wetness index, aspect, and slope of the footprint modeled source area. **We use the variable dropping**

algorithm by Elith et al. (2008) to reach a compromise between predictive deviance and model parsimony. This algorithm (i) fits a BRT model, (ii) performs a 10-fold CV, (iii) drops the least important predictor (improvement to the model, number of splits, Friedman, 2001), and (iv) repeats this sequence until a stopping criterion is reached. The mean CV deviance can be used to decide how many variables can be removed without significantly affecting predictive performance. Here, we set an upper threshold of 30 W m^{-2} for the mean CV deviance, which equals $\leq 1/2$ the random sampling error in the flux observations (Table 4). The dropping of variables first stopped for LE at 29.2 W m^{-2} mean CV deviance, yielding a set of the five most important predictors (LST, EVI, S_{\downarrow} , MR, and θ). For H the same predictor set yields a mean CV deviance of only 22.6 W m^{-2} . Remarkably, atmospheric pressure, z , and z_i were no significant predictors for the observed fluxes. This indicates that the chosen flight/analysis strategy effectively minimizes cross-contamination of the flux observations by vertical flux/pressure gradients. Analogously the algorithm dropped elevation, aspect, and slope of the footprint modeled source area as predictors. This shows that slope-induced effects on radiative transfer or turbulence generation do not significantly impact the flux observations. Consequently, the final BRT model is fitting an ERF to H and LE as function of only the five most important predictors. This ERF is then used to predict H and LE throughout the XRC, as a function of LST and EVI for each grid cell, and the median S_{\downarrow} , MR, and θ for the duration of a flight.

In the following we will use the term LTFM, referring to the overall procedure consisting of Low level flights, Time-frequency-, Footprint-, and Machine learning analyses (Fig. 5).

2.5 Uncertainty

Throughout the present study, we use the median and the median absolute deviation as preferred measure of location and scale, respectively (Croux and Rousseeuw, 1992; Rousseeuw and Verboven, 2002). All resulting uncertainty estimates are representative of one standard deviation.

For the purpose of detecting systematic differences between observations and predictions, we use the maximum-likelihood fitting of a functional relationship (MLFR, Ripley and Thompson, 1987). This method assigns a weight to each data couple in the relationship, which is inversely proportional to its error variances. In our case, the squared random flux errors in the observa-

tions, and the residuals in the BRT cross validation ensemble are used. This appreciates reliable data and depreciates uncertain data couples. The errors in the MLFR coefficients are determined from a jack-knife estimator (Quenouille, 1956; Tukey, 1958). If the regression intercepts were not significant, the relationships were forced through the origin, and confidence intervals were determined from the slope error. The coefficient of determination R^2 was calculated in analogy to weighted least-squares regression (Kvalseth, 1985; Willett and Singer, 1988). It is the proportion of variation in the weighted dependent variable that can be accounted for by the weighted independent variable.

Uncertainty in the LTFM up-scaling procedure originates from different sources during measurement and data analysis. Part of these uncertainty terms exhibit random characteristics, i.e. they tend to cease with sample size. Another part however will systematically bias the results, independent of sample size. An uncertainty budget for the random and systematic uncertainties in the LTFM procedure will consist of uncertainty terms for (i) instrumentation and hardware, (ii) turbulence sampling, (iii) spatio-temporal analysis, (iv) BRT residuals, (v) BRT response function, and (vi) BRT state variables. While uncertainty terms (i), (ii), and (iv) can be quantified with readily available procedures (Sects. 2.4, 3.3), we will describe several techniques to assess terms (iii), (v) and (vi) in the following.

2.5.1 Spatio-temporal analysis in heterogeneous terrain

Under the umbrella of spatio-temporal analysis we quantify in the following the uncertainty contribution from Wavelet analysis, footprint modelling, and the assumption of linear mixing. The fluxes derived from the wavelet cross-scalogram were adjusted to match the leg-averaged fluxes from time-series EC, which avoids bias between both techniques. Also areas above the wavelet cross-scalogram COI were used in the flux calculation to ensure including all scales of turbulent transport along the entire transect. However, values above the COI are potentially distorted due to edge effects, in particular close to the beginning and the end of each transect. These artefacts propagate in the resulting variances and fluxes, and consequently into the footprint estimates. Additional spatial uncertainty terms result from the use of an “offline” footprint model that does not consider the actual flow field, as well as from the MODIS EVI and LST

Discussion Paper | Discussion Paper | Discussion Paper

data. The use of BRT does not expect a linear response between the state variables and the flux signal. However, LTFM still assumes the linear mixing of the flux signal with respect to the contributing surface patches with different biophysical properties and source strengths.

To quantify the error inherent in the above analysis steps, we compare maps of LTFM predicted fluxes to airborne flux observations. (a) The BRT is trained with all available observations ($N = 8466$), (b) using the median state variables along each flight leg ($N = 42$), the BRT response function is used to predict a similar number of flux maps (Fig. 11), (c) the LTFM footprints are superimposed over these flux maps, (d) for each flux observation a predicted flux is calculated as the footprint-weighted average of all contributing cells, and (e) predictions and observations are compared.

2.5.2 Response function

BRTs are a non-parametric machine learning technique in which a response function is constructed according to the coherencies in the training data. As a direct consequence the predictive performance of BRT depends on how complete the combinations of state variables in the evaluation data are represented in the training data. Here we assess the susceptibility of the BRT response function and predictive performance to missing state variable combinations in the training data. For this purpose, 12 incomplete training data sets are created, each of which omitting a different flight out of the total of 12 flights in Table 1. For each incomplete training data set, (a) the BRT is trained, (b) the resulting response function is used with the state variables along the omitted flight for prediction, and (c) predictions and observations are compared.

2.5.3 State variables

Here, we consider the uncertainty resulting from disregarding part of the natural variability in the state variables that are used for spatially and temporally explicit BRT predictions. For this purpose we quantify the disregarded parts of the natural variability in each state variable and propagate it through the full BRT model. While explicit in time, the meteorological variables measured by the aircraft do not cover the entire catchment. We estimate a measure of spatial

variability from all subsequent pairs of flights that are located in different areas of the catchment (Table 1, Fig. 2). The median differences throughout the catchment for $S\downarrow$ ($-6 \pm 12 \text{ W m}^{-2}$), θ ($-1.1 \pm 1.1 \text{ K}$), and MR ($-0.5 \pm 0.3 \text{ K}$) are not significant (Wilcoxon rank-sum test, $p \geq 0.18$). On the contrary, MODIS EVI and LST are explicit in space, but not continuous in time. The 8-day trends from one scene to the next are accounted for in the BRT procedure through temporal interpolation between the MODIS scenes (Sect. 2.2). However, processes of shorter duration, such as frequent events of small-scale convective precipitation, go unaccounted. Hence we estimate a measure of the natural variability between two MODIS scenes. For this purpose we calculate the median change of all grid cells between all subsequent MODIS scenes, amounting to 0.01 ± 0.05 for EVI and $-0.5 \pm 6.2 \text{ K}$ for LST. The random part of EVI and LST natural variability by far exceeds the MODIS data product uncertainty of ≈ 0.015 (Xiang et al., 2003) and $\approx 1 \text{ K}$ (Wan and Li, 2008), respectively. Hence MODIS data product uncertainty was not considered separately.

The correlation matrix between the state variables was calculated using all 8446 aircraft observations. A variance-covariance matrix was calculated from this correlation matrix and the random part of the state variables' natural variability. Preserving the variance-covariance relationship, 1000 samples were drawn from a multivariate normal distribution with zero mean. These represent 1000 combinations of co-existing natural variability in the state space of the BRT model. The propagation through the BRT model was performed individually for each combination by (a) superimposing the estimated natural variability over the measured state variables of all 8446 observations, (b) performing a BRT prediction, and (c) comparing the results to the undisturbed predictions.

3 Results and discussion

In the first part of this section we assess the surface-atmosphere mixing regimes. From there wavelet analysis, footprint modelling and BRTs are used to infer ERFs between land surface properties and the flux measurements. Lastly, uncertainties in the LTFM up-scaling procedure are analysed and discussed.

3.1 Horizontal mixing between surface and flight level

On spatial average energy conservation requires that the vertical profiles of H and LE approach their respective entrainment flux at the top of the CBL (e.g. Deardorff, 1974; Sorbjan, 2006). The linear vertical flux gradient of H throughout the CBL was calculated ($-0.21 \text{ W m}^{-2} \text{ m}^{-1} - 0.06 \text{ W m}^{-2} \text{ m}^{-1}$), assuming that H ceases at the statically stable entrainment zone around 0.8 CBL. However, the entrainment flux of E is unknown. Hence we cannot estimate the vertical flux gradient of E , but assume a comparable order of magnitude as for H . The resulting effect of the vertical flux gradient below the flight level is $-5 \pm 2 \%$ of H , which falls well within the surface layer definition (e.g. Raupach and Finnigan, 1995; Stull, 1988, flux constant within $|5-10\%|$). Thus, it is feasible to assume that H and E measured at flight level are representative of surface fluxes.

The characteristic length scale of surface heterogeneity is in the order of several hundred to thousand meters, with an average of $L_H = 1012 \pm 715 \text{ m}$ (Table 2). Along identical flight paths L_H is comparable between days with different meteorological settings (e.g. 15 and 17 July 2009, 26 and 30 July 2009). This confirms the usefulness of the surface temperature measurement as proxy for surface heterogeneity. Only the longer flight paths C1, C2 cross the dune belt in the centre of the catchment (Fig. 2). The dune belt is the largest continuous land cover following steppe, and consequently the autocorrelation function of T_s estimates large values of L_H (1458–2615 m). During all flights, L_H was small compared to the Raupach length ($L_R = 1532-5214 \text{ m}$), and thus the influence of the surface heterogeneity is confined within the CBL ($z_i = 1100-2500 \text{ m}$). Here we use the thermal blending height ($z_{\text{TB1}} = 40 \pm 29 \text{ m}$) as an estimate for the vertical level where quasi-equilibrium of the turbulent exchange between land surface and atmosphere is reached. At all times the flight level ($z = 48-102 \text{ m}$) is above z_{TB1} and below $\approx 10 \%$ of the CBL depth, a common estimate for the depth of the atmospheric surface layer (e.g. Raupach and Finnigan, 1995; Stull, 1988). Hence it is feasible to assume that the turbulence measurement at flight level is representative for the land surface in the flux footprint. **The interpretation of the flux observations might be more complicated for measurement heights below the thermal blending height (limited spatial representativeness) or above**

the surface layer (vertical flux gradient). The blending length formulations L_{TB1} (1660 ± 723), L_{TB2} (957 ± 441) are used to assess the minimum size of surface heterogeneity that significantly influences the flow at flight level. Here we use $255 \text{ m} < L_{TB2} < 1852 \text{ m}$ as a guideline, because L_{TB2} is also representative of the magnitude of surface heterogeneity. The native resolution of the EVI data (230 m) and the land cover data (90 m) is equal or better than L_{TB2} , and thus sufficient to reproduce the variability of the land cover. In comparison, the native resolution of the LST data (1000 m) is coarse, potentially leading to an attenuation of the ERFs.

Using the wavelet cross-scalogram, long wavelength contributions to the flux do not constrain the spatial resolution of the flux computation along the flight path. Nevertheless, the random flux error is inversely proportional to the square root of the averaging length (e.g. Lenschow and Stankov, 1986), and propagates directly into the computation of the ERFs. Hence, we consider a trade-off between random error (high resolution) and smearing (low resolution) of the resulting flux estimates. The upwind distance (perpendicular to the WSMA flight path) where 80 % of the flux contributions are included in the footprint, $L_{80\%} = 1171 \pm 314 \text{ m}$, is comparable in magnitude to L_H . Thus, a flight path length of similar extent (1000 m) is a physically meaningful window for the computation of turbulence statistics and fluxes because (i) changes in the turbulent flux (response) are resolved at the same spatial scale as the characteristic surface heterogeneity (driver); (ii) the turbulence statistics used for footprint calculations are representative on the same spatial scale as the upwind extent; and (iii) the random error for each flux estimate decreases by $\approx -70 \%$ compared to a window length of 90 m.

The (aerodynamic) roughness length is usually below 1 m, with exception of the low wind speed situation on 17 July 2009, pattern O11, and the higher flight levels ($z \geq 97 \text{ m}$) on 26 July 2009 (Table 2).

3.2 Flux un-mixing

The presentation of the flux un-mixing results follows the sequence of LTFM analysis steps. In Sect. 3.2.1 the spatially resolved flux observations from the wavelet cross-scalogram are illustrated. Subsequently, footprint modelling is used to infer the biophysical surface properties in the source area of each flux observation (Sect. 3.2.2). In Sect. 3.2.3 the ERF between flux

observations and meteorological and land surface drivers is established. This ERF is then used to predict the surface fluxes throughout the XRC, which are finally summarized for different land covers (Sect. 3.2.4).

3.2.1 Spatially resolved flux measurement

5 Here and in the following we use a flight along pattern O12 for illustration, which follows a shallow elevation gradient (Fig. 4 bottom panel). This flight pattern is particularly suitable for this purpose because of its marked land cover changes over a relatively short distance. The wavelet cross-scalogram allows a high spatial discretization of turbulent flux measurements. **At the same time it includes flux contributions from wavelength that are significantly longer than**
10 **the 1000 m subinterval for each flux observation** (Fig. 4). The resulting high number of flux observations along a flight line leads to previously unachievable resolution and coverage of the state space. Spatially coherent flux contributions are detected on transport scales (eddy sizes) of 500–2000 m, that is of similar size as the extent of homogeneous surface patches L_H . Strong local flux contributions are confined to scales < 500 m, and approximately decay within the lower
15 threshold of the observed blending lengths L_{TB1} , L_{TB2} . This confirms a close coupling between atmospheric turbulence structures with surface patchiness, and consolidates the interpretation of the length scale approach. The less certain flux contributions above the wavelet COI are small (-15 to -4 % median differences for all flights). In the present example the COI is confined to relatively small scales (≤ 4 km), which is a direct result of the comparatively short flight. In
20 general, more certain flux contributions below the COI include transport scales up to $\approx 1/3$ of the flight length, and can reach ≈ 16 km for flight patterns C1 and C2. However, this also implies that the maximum considered transport scale differs between the flight patterns, just as it would be the case for the time-domain EC method. The wavelet cross-scalogram reveals strong turbulent transport in the second and fourth quarter of the flight for H , and in the first and third
25 quarter for LE (Fig. 4). When integrated over all transport scales for each overflowed 90 m cell of the land cover grid, these patterns correspond to strong upward fluxes.

3.2.2 Land cover

In Sect. 3.2.1, turbulence statistics and fluxes were integrated for each overflown 90 m cell of the land cover grid. In the following we expand the integration window to overlapping subintervals of 1000 m length, while retaining a spatial discretisation of 90 m. Such procedure significantly reduces the random sampling uncertainty (Sect. 3.1), though at the cost of decreasing the number of resulting observations by one window size ($dN \approx 10$). The resulting turbulence statistics are used to calculate the source area of each individual flux observation along the flight line, which are superimposed over the land cover grids. Figure 6 shows that in general LST and EVI follow the land cover patterns, e.g. lower temperature and higher greenness for irrigated agriculture and marshland. However, it is also evident that the static land cover classification cannot reflect the current surface conditions. E.g. the marshland in the north-western quadrant appears dried-out (high LST and low EVI), while the steppe area in the North-Eastern quadrant shows large variations in LST. Hence, biophysical surface properties also vary significantly within the land cover classes. This is likely a function of geomorphological properties such as aspect, slope and soil type, but also due to the large variability of convective rainfall events across the study area (e.g. Schaffrath et al., 2011).

Following superimposition of the footprints over the land cover data, the spatial contributions of different surface properties to each flux observation can be quantified (Fig. 7). It is evident that measured Bo changes in correspondence with the dominating land cover, i.e. low Bo for marshland and irrigated agriculture, and high Bo for bare soil and steppe. LST and EVI are stratified between the land covers, although in different sequence compared to the regional average (Fig. 3). The variability of LST and EVI within the land cover classes is equal or larger compared to the between-class variability, in particular for marshland, irrigated- and rainfed agriculture. While LST and EVI behave inversely for all natural land covers ($-0.78 < r < -0.10$), the contrary is true for irrigated- ($r = 0.92$) and rainfed ($r = 0.30$) agriculture. The latter finding appears counter-intuitive, but can be explained by tillage farming in the low-level plains with crops that are not adapted to the semiarid climate, such as potatoes. The albedo of these densely vegetated crops can be lower compared to the sparsely vegetated steppe land

cover ($\alpha \approx 0.2$, Ketzer et al., 2008), resulting in higher foliage temperatures. Only two natural land covers, marshland and mountain meadow exhibit similarly high EVI values as the field crops (Figs. 2, 3). Nevertheless the LST of these land covers is comparatively low. In case of the marshland this can be explained by water saturated soils with high heat capacity. Conversely, lower temperatures in accordance with the adiabatic temperature gradient are expected for the mountain meadows at higher altitudes.

In Fig. 8 H and LE observations along the flight line are shown together with the LST and EVI in the respective source area. Because of the 1000 m integration window over the wavelet cross-scalogram the results appear smoother compared to Fig. 4, where a 90 m integration window is used. It is apparent that H and LE both systematically change with LST ($r_H = 0.64$, $r_{LE} = -0.84$) and EVI ($r_H = -0.62$, $r_{LE} = 0.73$). However, peaks in H (3 km and 9 km in Fig. 8) and in LE (0 km and 7 km) don't manifest when their respective land surface drivers in the footprint are maximal. Instead they seem to follow a trade-off function between LST and EVI.

3.2.3 Environmental response functions

Thus far our findings indicate that the interactions between land surface and atmosphere are multi-faceted and potentially non-linear. Hence we use LST and EVI as topical, spatio-temporal proxies for the source strength of H and LE, rather than using the land cover classification directly. In comparison to earlier flux un-mixing studies (Chen et al., 1999; Hutjes et al., 2010; Ogunjemiyo et al., 2003) this has the benefit of (i) providing individual source strength representations for the effects of surface moisture and temperature, and (ii) representing the land surface by continuous (LST, EVI) rather than discrete variables (land cover classes), thus enabling the use of more advanced scaling algorithms.

Here, we use BRTs to extract the relationships between all ($N = 8446$) flux observations and land cover (LST, EVI) and meteorological (S_{\downarrow} , MR, and θ) variables. While BRTs are capable of reproducing complex interactions through multi-layered branching, the fitted function can be summarized, e.g. as partial dependence plots (Fig. 9). These show the effect of each individual variable on the response after (i) subtraction of the offset ($H_0 = 161 \text{ W m}^{-2}$,

$LE_0 = 176 \text{ W m}^{-2}$), and (ii) accounting for the average effects of all other variables in the model. The partial dependence plots in Fig. 9 are sorted in order of the relative importance of the response variables (Friedman, 2001). The most important responses of H are non-linear (LST, θ), followed by linear responses ($S\downarrow$, MR, and EVI). With the exception of MR and EVI the individual responses are positive in sign. The order of the responses for LE is partially different (MR, LST, θ , $S\downarrow$, and EVI), and only the responses on $S\downarrow$ and EVI are approximately linear (not shown). With exception of MR (concave, maximal response around 10 g/kg) and LST (convex, minimal response around $\approx 310 \text{ K}$), the signs of the responses for LE are positive. It appears surprising that H and LE are only weakly related to $S\downarrow$. This can be explained by using only noontime flights in the present study, where $S\downarrow$ mainly fulfils the purpose of accounting for varying cloud/radiation conditions between different measurement days. In addition, during individual flights $S\downarrow$ was usually constant to within $\leq 10\%$ (Table 5). When however using ERFs to reproduce a diurnal cycle, a much larger dependence of H and LE on $S\downarrow$ would be expected.

In Fig. 10 MLFRs are established between BRT fitted values for H and LE and the observed fluxes ($N = 8446$). Here we use the BRT cross-validation residuals and the random sampling errors in the observations to determine the MLFR weights of each data point. Uncertainty terms (i), (iii), (v) and (vi, Sect. 2.5) cannot be quantified individually for each observation. Hence these terms are not considered here, but in the final uncertainty budget (Tables 3 and 4). For both, H and LE the agreement between the BRT fitted values and the observed fluxes is excellent. Contrary to our initial anticipation, the ERFs are not attenuated by the relatively coarse MODIS LST resolution, as indicated by approximately zero MLFR offset and unity slope. The median absolute deviation in the residuals is small ($\leq 1\%$). However, several outliers are found for moderate to high fluxes of H ($N = 41$) and LE ($N = 133$), for which the BRTs underestimate the observed value by -150 W m^{-2} or more. The majority of these cases occur during the flights O8 on 13 July 2009 and C1 on 26 July 2009, respectively. On both dates the outliers concur with highly intermittent solar irradiance ($200 < S\downarrow < 1200 \text{ W m}^{-2}$) along a short section of the flight paths. For instance an intermittent cloud cover can disrupt the functional relation between the irradiance (driver) and the flux (response) observations because

(i) at a flight level of 50–100 m a.g.l., the aircraft irradiance measurement does not represent S_{\downarrow} in the source areas of H and LE, and (ii) the plant physiological response can vary substantially on spatio-temporal scales that are small compared to atmospheric transport processes between the land surface and the aircraft. Our choice of land surface and meteorological drivers appears to work well for describing the noontime surface-atmosphere exchange of heat and water vapour over a moisture-limited landscape. However, it is important to note that appropriately describing exchange processes over longer periods of time, for different landscapes or scalars might require finding an entirely different set of predictors.

3.2.4 Extrapolation and summarization

For the duration of each flight pattern, the trained BRT models are used to extrapolate H and LE throughout the XRC. For this purpose the median meteorological state variables during each flight pattern as well as topical grids of MODIS LST and EVI data are used. Grid cells that exceed the state space of the BRT training data set ($N = 8446$) are excluded from extrapolation. Fig. 11 shows the resulting flux grids for three different days, with a spatial coverage of $\geq 92\%$. Because of the identical state space ranges for BRT training and prediction, also the ranges of the extrapolated turbulent fluxes are within limits of the observations. Despite the land cover classification was never used during the extrapolation process, several landscape units are clearly recognizable in the flux maps. For instance both, the Xilin river valley, as well as the mountainous headwater area to the east display low H and LE. On the contrary, the non-vegetated basin on the Northern tip shows consistently low evapotranspiration.

For a given meteorological boundary condition (MR, θ , S_{\downarrow}), the heat fluxes within several hours of solar zenith can be expressed as a function of LST and EVI (Fig. 9). In turn, these biophysical surface properties are characteristic within a land cover class (Fig. 3). Here, we aggregate all grid cells of the flux maps according to land cover class, resulting in sample distributions of H and LE. This allows a formal transition from a mosaic- to a tile representation of H and LE over the XRC for the duration of each flight pattern (Mengelkamp et al., 2006). These sample distributions then enable the analysis of land cover specific source strengths ($36 \text{ W m}^{-2} < H < 364 \text{ W m}^{-2}$, $46 \text{ W m}^{-2} < \text{LE} < 425 \text{ W m}^{-2}$), as well as the spatial variability

ity within a land cover ($11 \text{ W m}^{-2} < \sigma_H < 169 \text{ W m}^{-2}$, $14 \text{ W m}^{-2} < \sigma_{LE} < 152 \text{ W m}^{-2}$). Table 3 gives an overview of the median land cover specific H and LE over all flight patterns, and their median spatial variability. These results fall well in the range of summertime ensemble average fluxes during solar noon observed by ground-based EC measurements over different land covers in this region ($100 \text{ W m}^{-2} < H < 310 \text{ W m}^{-2}$ and $100 \text{ W m}^{-2} < LE < 480 \text{ W m}^{-2}$, Gao et al., 2009; Hao et al., 2007; Hao et al., 2008; Shao et al., 2008). In comparison, the flight-line average heat fluxes are in the range of $71 \text{ W m}^{-2} < H < 310 \text{ W m}^{-2}$ and $46 \text{ W m}^{-2} < LE < 300 \text{ W m}^{-2}$ (Table 6).

However, the magnitudes of H and LE are not only functions of land cover, but also proportional to the available energy. The available energy changes within, but in particular between flight days (Table 1). To alleviate this effect and to enable the comparison between different flights, we calculate the Bowen ratio $Bo = H/LE$ between the sample distributions. Despite differences in the meteorological drivers (MR, θ , $S\downarrow$) the median land cover specific Bo agrees well between subsequent flight patterns on all measurement days (Fig. 12). During the afternoon flights $12 \pm 9\%$ higher Bo values are observed compared to the morning flights, as expected from a land surface that desiccates in the course of the day. Nevertheless the 99.9% confidence interval includes unity slope. Hence, for several hours within solar zenith Bo does not change significantly, and can be interpreted as a characteristic land surface property. On this basis we summarize the regional flux estimates for the duration of the flight campaign as time series of land cover specific Bo ratios (Fig. 12). The order of Bo between the land covers largely follows the order of the land cover specific EVI, while the temporal pattern follows the pattern of the land cover specific LST (Fig. 3). High Bo values until mid-campaign indicate that the land surface dries out. This trend is reversed toward the end of the campaign, when the approach of humid air masses leads to considerable precipitation. The median daily natural variability of Bo within the land covers ranges from 48% (rainfed agriculture) to 79% (marshland). Water absorbs strongly in the near infrared, leading to negative EVI values that are not indicative of vegetation greenness. Hence EVI values for water surfaces are discarded, and the land cover “Water” cannot be modeled by the present ERF.

3.3 Uncertainty

Metzger et al. (2012) have shown that turbulent flux measurements with the WSMA platform and instrumentation are unbiased, and precise to within 8 %. Uncertainty due to the limited sampling size of turbulent eddies is estimated using the methods of Lenschow and Stankov (1986);
5 Lenschow et al. (1994). Details on the implementation can be found in Metzger et al. (2012). For a single flux measurement the systematic (and random) components of this sampling uncertainty range from $< 1 \%$ (57 %) for H to $< 1 \%$ (121 %) for LE. Table 4 summarizes above uncertainty sources, as well as additional sources which are treated in the following.

In order to assess the uncertainty arising from the spatio-temporal analyses (Sect. 2.5.1), we
10 compare the median observed and predicted fluxes along all flight legs (Fig. 13). The LTFM predictions slightly overestimate the observed fluxes ($H = 5 \%$, $LE = 5 \%$), but in both cases the 99.9 % confidence intervals include unity slope. The median differences of $dH = 2 \%$ (40 %), and $dLE = 4 \%$ (47 %) agree marginally closer. Moreover, the median residuals between fitted and observed values emphasize that the BRT fitting technique is unbiased (Table 4).

15 Subsequently, we assess the predictive performance of the BRT response function in the light of missing state variable combinations in the training data. For this purpose one flight at a time was omitted from the training data, and the incompletely trained BRT model was used to predict the missing data. The resulting median differences amount to 11 % (69 %, $N = 7311$) for H and 18 % (77 %, $N = 7265$) for LE. During prediction, cases where one or more state variables
20 exceed their respective range during training were excluded. As a consequence the sample size is $\approx 14 \%$ smaller than the total number of observations ($N = 8466$).

Lastly, we consider the uncertainty resulting from disregarding part of the spatio-temporal variability in the state variables during BRT predictions. For this purpose we quantify the disregarded parts of the natural variability, and propagate it through the full BRT model. The result-
25 ing median differences amount to 13 % (77 %), and 14 % (75 %) for H , and LE, respectively, and are dominated by the effect of LST natural variability ($r = 0.81$, and -0.69). Because the response of the BRT predictions on LST is non-linear (Fig. 9), deviations of similar magnitude but opposite sign in LST do not cancel out in the predictions. This can lead to a systematic

overestimation as a function of the specific state variable combination in each prediction, and is hence dependent on the catchment composition. However, in all test cases the 99.9% confidence intervals between observed and predicted fluxes include unity slope. Hence we go without introducing a non-linearity response factor, but assign an accuracy of $\leq 20\%$ to the LTFM method.

Assuming normal distribution and independence, the random parts of all uncertainty terms (in parenthesis, Table 4) can be combined to their Gaussian sum. Then, the ensemble random uncertainty σ_{ens} considers the reduction of the random uncertainty with sample size (e.g. Mahrt, 1998);

$$\sigma_{\text{ens}} = \frac{\sigma_{\text{ran}}}{\sqrt{N}}, \quad (10)$$

with zero expected value σ_{ens} and the SD σ_{ran} of the population with size N . While σ_{ran} is a measure for the average dispersion of a single observation or grid cell, σ_{ens} quantifies the level of confidence we can expect from aggregating multiple observations or grid cells. The resulting ensemble random uncertainty for land cover specific flux estimates throughout the XRC ranges from $< 1\%$ for steppe to 5% for settlements and irrigated agriculture (Table 3).

4 Conclusions

The overarching goal of airborne EC flux measurements is to bridge the gap between observations and data assimilation approaches on different spatial scales. This study develops the LTFM procedure to characterize the exchange of sensible and latent heat for different land covers in a heterogeneous steppe landscape. The procedure “mines” the information content of EC flux observations and extracts quantitative relationships with environmental drivers. In the process LTFM maximises objectivity and data use efficiency – all available observations are considered. The subsequent steps of LTFM are (i) low level EC flux flights, (ii) time-frequency analysis of the flux observations, (iii) source area modelling of continuous biophysical surface properties, and (iv) inferring ERFs from non-parametric machine learning.

1. The use of a weight-shift microlight aircraft with low airspeed and high climb rate enables low level flights at constant height even above topographically structured terrain. Masking out slopes during flight planning effectively minimizes cross-contamination of the flux observations by slope-induced effects on radiative transfer or turbulence generation. This reduces the degrees of freedom in explaining the observed flux responses, albeit potentially at the expense of oversimplifying surface-air exchange processes.
2. Wavelet decomposition of the turbulence data yields unprecedented spatial resolution of the flux observations. However, due to edge effects flux observations close to the start or end of a dataset can contain spectral artefacts. Using alternative techniques such as empirical mode decomposition (Barnhart et al., 2012a,b) or structure-parameter methods (Van Kesteren et al., 2013) might help to further improve the results.
3. An “offline” footprint parameterization considering 3-D dispersion is suitable to map the differences in surface properties encountered by a flux aircraft. However, when adapting LTFM, e.g. to ground-based measurements, the range of surface properties is likely to shrink significantly. In order to improve the deteriorated signal to noise ratio it might become important to also consider the local flow field, especially when measuring at greater heights. E.g. closure models with terrain-following coordinates (Hsieh and Katul, 2009; Sogachev and Lloyd, 2004) or “online” Lagrangian dispersion modelling (Markkanen et al., 2010; Matross et al., 2011; Wang and Rotach, 2010; Weil et al., 2012) could be useful for such purpose.
4. In spite of a static and discrete land cover classification, the LTFM method uses spatio-temporally continuous and topical information of biophysical surface properties. Only the continuous nature of MODIS land surface data enabled the use of the BRT machine learning technique. In this combination the climatic and altitudinal gradients throughout the XRC are successfully reproduced. In the interest of further advancing LTFM, it is desirable to also consider the uncertainty in the observations during machine learning, and to explore alternative machine learning techniques such as support vector machines (e.g. Yang et al., 2007).

The ERFs resulting from LTFM can aid bridging observational scales by, e.g. isolating and quantifying relevant land-atmosphere exchange processes, estimating land cover specific emission factors, extending flux measurements to the catchment scale, assessing the spatial representativeness of EC flux measurements etc. Analogously applying LTFM to ground-based EC measurements could aid, e.g. advancing the treatment of location bias from diagnostic assessment (e.g. Chen et al., 2012) to prognostic transfer functions, constraining local to regional water budgets, distinguishing anthropogenic and natural sources/sinks in urban environments and substantiating process-studies.

Appendix A Notation

A1 Abbreviations

3-D	Three-dimensional
a.g.l.	Above ground level
a.s.l.	Above sea level
Arable	Rainfed agriculture
ASTER	Advanced Spaceborne Thermal and Reflection Radiometer
Bare	Bare soil
BRT	Boosted regression trees
CBL	Convective boundary layer
COI	Cone of influence
CST	Chinese standard time (CST = Coordinated universal time + 8)
CV	Cross-validation
DFG	German Research Foundation
EC	Eddy covariance
ERF	Environmental response function
IMGERS	Inner Mongolia Grassland Ecosystem Research Station
Irrigated	Irrigated agriculture

KL04	Footprint parameterisation of Kljun et al. (2004)
KL04+	Footprint parameterisation of Kljun et al. (2004) with superimposed cross-wind dispersion function
LTFM	Low level flights, Time-frequency-, Footprint-, and Machine learning analyses
MAGIM	Matter fluxes in grasslands of Inner Mongolia as influenced by stocking rate
Marsh	Marshland
MLFR	Maximum likelihood functional relationship
MODIS	Moderate Resolution Imaging Spectroradiometer
Mountain	Mountain meadow
SD	Standard deviation
Settle	Settlements
Steppe	Generic steppe
WSMA	Weight-shift microlight aircraft
XRC	Xilin River Catchment

A2 Functions

Overbars denote the mean along a flight line, and primes denote the deviations from this mean.

*	Complex conjugate
cov	Covariance
d	Difference
σ	Standard deviation
ψ	Mother wavelet

A3 Parameters and variables

α	Albedo (–)
----------	------------

a	Wavelet scale parameter (s)
a_0	Initial wavelet scale parameter (s)
b	Wavelet location parameter (s)
Bo	Bowen ratio (–)
c	Lag of autocorrelation function (m)
CC	Cloud cover (–)
C_δ	Wavelet reconstruction factor (–)
d	Distance along a flight line (m)
DIR	Wind direction ($^\circ$)
δj	Wavelet frequency increment (–)
δt	Wavelet time increment (s)
e	Euler's number ≈ 2.71828 (–)
EVI	Enhanced vegetation index (–)
H	Sensible heat flux (W m^{-2})
i	Imaginary unit $i^2 = -1$ (–)
ID	Flight identifier (–)
j	Running index (–)
J	Number of wavelet scale increments (–)
l	Length of flight line (km)
L	Monin-Obukhov length (m)
$L_{80\%}$	Upwind distance where 80 % of the flux contributions are included in the footprint (m)
LC	Land cover class coverage (%)
LE	Latent heat flux (W m^{-2})
LST	Land surface temperature (K)
L_H	Horizontal scale of surface heterogeneity (m)
L_R	Raupach length (m)
L_{TB1}	Thermal blending length (m)
L_{TB2}	Improved thermal blending length (m)

MR	Mixing ratio (gkg^{-1})
n	Running index (-)
N	Sample size (-)
p	Probability of test statistic (-)
P	Cumulated precipitation in a 10 day trailing window (mm)
q	Dimensionless wavelet coordinate (-)
r	Pearson correlation coefficient (-)
rep	Repetitions (-)
S_{\downarrow}	Down-welling shortwave radiation (Wm^{-2})
t	Time (s)
T_s	Surface temperature (K)
θ	Potential temperature (K)
$\theta_{0,v}$	Virtual potential temperature (K)
u	Horizontal wind speed (ms^{-1})
u_*	Friction velocity (ms^{-1})
w	Footprint weight (-)
W	Wavelet coefficient (-)
w_*	Convective velocity (ms^{-1})
x	Wildcard for a signal (-)
y	Wildcard for a signal (-)
z	Measurement height (m)
z_0	Aerodynamic roughness length (m)
ω_0	Wavelet frequency parameter (-)
z_i	Convective boundary layer depth (m)
z_{TB1}	Thermal blending height (m)

A4 Subscripts

In general, subscripts follow the parameter and variable definitions in Appendix A3. Instances with differing use of subscripts are defined in the following.

ens	Ensemble
ran	Random
v	Cross-wind component
w	Vertical wind component
x	Longitudinal coordinate
y	Latitudinal coordinate

Appendix B Meteorological conditions

5 The midday flights are usually accompanied by a thin layer of cirrus clouds, interspersed with local convective cumuli, resulting in a cloud cover between 4/8 and 7/8 (Table 5). The down-
 welling shortwave radiation is decreasing over the duration of the campaign, with minima on
 17 and 30 July 2009. These minima coincide with the advection of comparatively moist air, as
 evident from the higher mixing ratios. S_{\downarrow} also correlates with the precipitation history (Table 1,
 10 $r = -0.40$). The wind speed at flight level decays from up to 8.3 ms^{-1} at the beginning down to
 2.7 ms^{-1} towards the last quarter of the flight campaign. All wind sectors with the exception of
 northerlies occur. Both, the virtual potential air temperature and the surface temperature peak
 during the middle of the flight campaign. As a result of several convective precipitation events
 the mixing ratio increases over the flight campaign, accompanied by a dampening of the surface
 15 temperature variability.

The range of the flight line average turbulent fluxes are $0.3 < u_* < 0.5 \text{ ms}^{-1}$, $71 \text{ W m}^{-2} < H < 310 \text{ W m}^{-2}$ and $46 \text{ W m}^{-2} < LE < 300 \text{ W m}^{-2}$ (Table 6). The friction velocity peaks during flights under high wind speeds. While H dominates the heat exchange during the middle of the campaign, LE peaks at the beginning and end of the campaign. The Bowen ratio throughout

the campaign correlates ($r = -0.67$) with precipitation history (Table 1), i.e. the moisture available for evapotranspiration. Moreover H clearly correlates with $S\downarrow$ ($r = 0.68$), while no such relationship was found for LE ($r = 0.02$). The atmospheric stratification was unstable throughout all flights (Monin-Obukhov length $L_v = -34 \pm 20$ m), with corresponding high values of the SD of the vertical wind $\sigma_w = 0.88 \pm 0.11$ ms⁻¹ and convective velocity $w_* = 2.21 \pm 0.39$ ms⁻¹.

Acknowledgements. We wish to thank Henry Loescher and Jeffrey Taylor at the National Ecological Observatory Network, Fundamental Instrument Unit, for their continued support and statistical advice. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc. This material is based upon work supported by the National Science Foundation under the grant DBI-0752017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Thanks go to Matthias Reiche at the Leibniz Centre for Agricultural Landscape Research, Institute of Soil Landscape Research, who provided the ASTER imagery used in this study. Recognition has to be given to Benjamin Blank, formerly Justus-Liebig-University Giessen, Institute of Landscape Ecology and Resources Management, who prepared the overview Fig. 1. The MODIS data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Centre, USGS/Earth Resources Observation and Science Centre (https://lpdaac.usgs.gov/get_data). The flight campaign in Inner Mongolia was funded by the German Research Foundation, research group FOR 536 MAGIM, “Matter fluxes in grasslands of Inner Mongolia as influenced by stocking rate”, and the National Natural Science Foundation of China, grant number 41021004. Stipend funding by the German Academic Exchange Service, Helmholtz Association of German Research Centres, China Scholarship Council and the European Union under the Science and Technology Fellowship China is acknowledged. We thank Ankur Desai, Raymond Desjardins and Paul Stoy for their valuable comments in the review process.

The service charges for this open access publication have been covered by the Open Access Publishing Fund of the Karlsruhe Institute of Technology, a Research Centre of the Helmholtz Association, and by the German Research Foundation.

References

- Anderson, M. C., Kustas, W. P., Alfieri, J. G., Gao, F., Hain, C., Prueger, J. H., Evett, S., Colaizzi, P., Howell, T., and Chávez, J. L.: Mapping daily evapotranspiration at Landsat spatial scales during the BEAREX'08 field campaign, *Adv. Water Res.*, **50**, 162–177, doi:[10.1016/j.advwatres.2012.06.005](https://doi.org/10.1016/j.advwatres.2012.06.005), 2012.
- Auerswald, K., Wittmer, M. H. O. M., Männel, T. T., Bai, Y. F., Schäufele, R., and Schnyder, H.: Large regional-scale variation in C3/C4 distribution pattern of Inner Mongolia steppe is revealed by grazer wool carbon isotope composition, *Biogeosciences*, **6**, 795–805, doi:[10.5194/bg-6-795-2009](https://doi.org/10.5194/bg-6-795-2009), 2009.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U., K., Pilegaard, K., Schmid, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, **82**, 2415–2434, doi:[10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2), 2001.
- Bange, J., Spiess, T., Herold, M., Beyrich, F., and Hennemuth, B.: Turbulent fluxes from Helipod flights above quasi-homogeneous patches within the LITFASS area, *Bound.-Lay. Meteorol.*, **121**, 127–151, doi:[10.1007/s10546-006-9106-0](https://doi.org/10.1007/s10546-006-9106-0), 2006.
- Barnhart, B. L., Eichinger, W. E., and Prueger, J. H.: Introducing an Ogive method for discontinuous data, *Agr. Forest. Meteorol.*, **162–163**, 58–62, doi:[10.1016/j.agrformet.2012.04.003](https://doi.org/10.1016/j.agrformet.2012.04.003), 2012a.
- Barnhart, B. L., Eichinger, W. E., and Prueger, J. H.: A new eddy-covariance method using empirical mode decomposition, *Bound.-Lay. Meteorol.*, **145**, 369–382, doi:[10.1007/s10546-012-9741-6](https://doi.org/10.1007/s10546-012-9741-6), 2012b.
- Beyrich, F., Leps, J. P., Mauder, M., Bange, J., Foken, T., Huneke, S., Lohse, H., Ludi, A., Meijninger, W. M. L., Mironov, D., Weisensee, U., and Zittel, P.: Area-averaged surface fluxes over the LITFASS region based on eddy-covariance measurements, *Bound.-Lay. Meteorol.*, **121**, 33–65, doi:[10.1007/s10546-006-9052-x](https://doi.org/10.1007/s10546-006-9052-x), 2006.
- Businger, J. A., Wyngaard, J. C., Izumi, Y., and Bradley, E. F.: Flux-profile relationships in the atmospheric surface layer, *J. Atmos. Sci.*, **28**, 181–189, doi:[10.1175/1520-0469\(1971\)028<0181:FPRITA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2), 1971.
- Butterbach-Bahl, K., Kögel-Knabner, I., and Han, X.: Steppe ecosystems and climate and land-use changes – vulnerability, feedbacks and possibilities for adaptation, *Plant Soil*, **340**, 1–6, doi:[10.1007/s11104-010-0651-4](https://doi.org/10.1007/s11104-010-0651-4), 2011.
- Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Black, T. A., Arain, M. A., Barr, A. G.,

- Bourque, C. P. A., Flanagan, L. B., Lafleur, P. M., McCaughey, J. H., and Wofsy, S. C.: Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis, *Remote Sens. Environ.*, 124, 742–755, doi:[10.1016/j.rse.2012.06.007](https://doi.org/10.1016/j.rse.2012.06.007), 2012.
- 5 Chen, J. M., Leblanc, S. G., Cihlar, J., Desjardins, R. L., and MacPherson, J. I.: Extending aircraft- and tower-based CO₂ flux measurements to a boreal region using a Landsat thematic mapper land cover map, *J. Geophys. Res. Atmos.*, 104, 16859–16877, doi:[10.1029/1999JD900129](https://doi.org/10.1029/1999JD900129), 1999.
- Chen, Z. Z.: Topography and climate of the Xilin river basin, *Research on grassland ecosystem*, 3, 13–22, 1988.
- 10 Croux, C. and Rousseeuw, P. J.: Time-efficient algorithms for two highly robust estimators of scale, *Computation. Stat.*, 1, 411–428, 1992.
- Deardorff, J. W.: Three-dimensional numerical study of turbulence in an entraining mixed layer, *Bound.-Lay. Meteorol.*, 7, 199–226, doi:[10.1007/bf00227913](https://doi.org/10.1007/bf00227913), 1974.
- Desjardins, R. L., MacPherson, J. I., Schuepp, P. H., and Hayhoe, H. N.: Airborne flux measurements of CO₂, sensible, and latent heat over the Hudson Bay lowland, *J. Geophys. Res. Atmos.*, 99, 1551–1561, doi:[10.1029/93JD01296](https://doi.org/10.1029/93JD01296), 1994.
- Desjardins, R. L., Macpherson, J. I., Neumann, H., Den Hartog, G., and Schuepp, P. H.: Flux estimates of latent and sensible heat, carbon dioxide, and ozone using an aircraft-tower combination, *Atmos. Environ.*, 29, 3147–3158, doi:[10.1016/1352-2310\(95\)00007-L](https://doi.org/10.1016/1352-2310(95)00007-L), 1995.
- 20 Desjardins, R. L., MacPherson, J. I., Mahrt, L., Schuepp, P., Pattey, E., Neumann, H., Baldocchi, D., Wofsy, S., Fitzjarrald, D., McCaughey, H., and Joiner, D. W.: Scaling up flux measurements for the boreal forest using aircraft-tower combinations, *J. Geophys. Res. Atmos.*, 102, 29125–29133, doi:[10.1029/97JD00278](https://doi.org/10.1029/97JD00278), 1997.
- Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, *Journal of Animal Ecology*, 77, 802–813, doi:[10.1111/j.1365-2656.2008.01390.x](https://doi.org/10.1111/j.1365-2656.2008.01390.x), 2008.
- 25 Emeis, S., Schäfer, K., and Münkel, C.: Surface-based remote sensing of the mixing-layer height – a review, *Meteorol. Z.*, 17, 621–630, doi:[10.1127/0941-2948/2008/0312](https://doi.org/10.1127/0941-2948/2008/0312), 2008.
- Fan, L., Liu, S., Bernhofer, C., Liu, H., and Berger, F. H.: Regional land surface energy fluxes by satellite remote sensing in the Upper Xilin River Watershed (Inner Mongolia, China), *Theor. Appl. Climatol.*, 88, 231–245, doi:[10.1007/s00704-006-0241-9](https://doi.org/10.1007/s00704-006-0241-9), 2007.
- 30 Friedman, J. H.: Greedy function approximation: A gradient boosting machine, *Ann. Stat.*, 29, 1189–1232, doi:[10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451), 2001.
- Gao, Z., Lenschow, D. H., He, Z., Zhou, M.: Seasonal and diurnal variations in moisture, heat and CO₂

- fluxes over a typical steppe prairie in Inner Mongolia, China, *Hydrol. Earth Syst. Sci.*, 13, 987–998, doi:[10.5194/hess-13-987-2009](https://doi.org/10.5194/hess-13-987-2009), 2009.
- Gioli, B., Miglietta, F., De Martino, B., Hutjes, R. W. A., Dolman, H. A. J., Lindroth, A., Schumacher, M., Sanz, M. J., Manca, G., Peressotti, A., and Dumas, E. J.: Comparison between tower and aircraft-based eddy covariance fluxes in five European regions, *Agr. Forest. Meteorol.*, 127, 1–16, doi:[10.1016/j.agrformet.2004.08.004](https://doi.org/10.1016/j.agrformet.2004.08.004), 2004.
- Glenn, E. P., Huete, A. R., Nagler, P. L., and Nelson, S. G.: Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape, *Sensors*, 8, 2136–2160, doi:[10.3390/s8042136](https://doi.org/10.3390/s8042136), 2008.
- Hao, Y., Wang, Y., Huang, X., Cui, X., Zhou, X., Wang, S., Niu, H., and Jiang, G.: Seasonal and inter-annual variation in water vapor and energy exchange over a typical steppe in Inner Mongolia, China, *Agric. For. Meteorol.*, 146, 57–69, doi:[10.1016/j.agrformet.2007.05.005](https://doi.org/10.1016/j.agrformet.2007.05.005), 2007.
- Hao, Y., Wang, Y., Mei, X., Huang, X., Cui, X., Zhou, X., and Niu, H.: CO₂, H₂O and energy exchange of an Inner Mongolia steppe ecosystem during a dry and wet year, *Acta Oecologica*, 33, 133–143, doi:[10.1016/j.actao.2007.07.002](https://doi.org/10.1016/j.actao.2007.07.002), 2008.
- Helmis, C., Sgouros, G., Tombrou, M., Schäfer, K., Münkel, C., Bossioli, E., and Dandou, A.: A comparative study and evaluation of mixing-height estimation based on SODAR-RASS, ceilometer data and numerical model simulations, *Bound.-Lay. Meteorol.*, 145, 507–526, doi:[10.1007/s10546-012-9743-4](https://doi.org/10.1007/s10546-012-9743-4), 2012.
- Högström, U.: Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation, *Bound.-Lay. Meteorol.*, 42, 55–78, doi:[10.1007/BF00119875](https://doi.org/10.1007/BF00119875), 1988.
- Holmes, R. M.: Meso-scale effects of agriculture and a large prairie lake on the atmospheric boundary layer, *Agron. J.*, 62, 546–549, doi:[10.2134/agronj1970.00021962006200040037x](https://doi.org/10.2134/agronj1970.00021962006200040037x), 1970.
- Hsieh, C.-I. and Katul, G.: The Lagrangian stochastic model for estimating footprint and water vapor fluxes over inhomogeneous surfaces, *Int. J. Biometeorol.*, 53, 87–100, doi:[10.1007/s00484-008-0193-0](https://doi.org/10.1007/s00484-008-0193-0), 2009.
- Hu, F. S., Higuera, P. E., Walsh, J. E., Chapman, W. L., Duffy, P. A., Brubaker, L. B., and Chipman, M. L.: Tundra burning in Alaska: Linkages to climatic change and sea ice retreat, *J. Geophys. Res.*, 115, G04002, doi:[10.1029/2009jg001270](https://doi.org/10.1029/2009jg001270), 2010.
- Hutjes, R. W. A., Vellinga, O. S., Gioli, B., and Miglietta, F.: Dis-aggregation of air-borne flux measurements using footprint analysis, *Agr. Forest. Meteorol.*, 150, 966–983, doi:[10.1016/j.agrformet.2010.03.004](https://doi.org/10.1016/j.agrformet.2010.03.004), 2010.
- Jiang, S.: An introduction to the Inner Mongolia Grassland Ecosystem Research Station, Research on

- grassland ecosystem, 1, 1–11, 1985.
- 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](https://doi.org/10.1029/97JD02559), 1997.
- 5 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 network design, *Atmos. Chem. Phys.*, 12, 7867–7879, doi:[10.5194/acp-12-7867-2012](https://doi.org/10.5194/acp-12-7867-2012), 2012.
- Ketzer, B., Liu, H., and Bernhofer, C.: Surface characteristics of grasslands in Inner Mongolia as detected
10 by micrometeorological measurements, *Int. J. Biometeorol.*, 52, 563–574, doi:[10.1007/s00484-008-0148-5](https://doi.org/10.1007/s00484-008-0148-5), 2008.
- Kirby, S., Dobosy, R., Williamson, D., and Dumas, E.: An aircraft-based data analysis method for discerning individual fluxes in a heterogeneous agricultural landscape, *Agr. Forest. Meteorol.*, 148, 481–489, doi:[10.1016/j.agrformet.2007.10.011](https://doi.org/10.1016/j.agrformet.2007.10.011), 2008.
- 15 Kljun, N., Rotach, M. W., and Schmid, H. P.: A three-dimensional backward lagrangian footprint model for a wide range of boundary-layer stratifications, *Bound.-Lay. Meteorol.*, 103, 205–226, doi:[10.1023/A:1014556300021](https://doi.org/10.1023/A:1014556300021), 2002.
- Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple parameterisation for flux footprint predictions, *Bound.-Lay. Meteorol.*, 112, 503–523, 2004.
- 20 Kljun, N., Rotach, M. W., and Schmid, H. P.: A simple parameterisation for two-dimensional flux footprints, in preparation, 2013.
- Kvalseth, T. O.: Cautionary note about R^2 , *Am. Stat.*, 39, 279–285, 1985.
- Lenschow, D. H. and Stankov, B. B.: Length scales in the convective boundary layer, *J. Atmos. Sci.*, 43, 1198–1209, doi:[10.1175/1520-0469\(1986\)043<1198:LSITCB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2), 1986.
- 25 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](https://doi.org/10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2), 1994.
- Liang, E., Shao, X., Hu, Y., and Lin, J.: Dendroclimatic evaluation of climate-growth relationships of Meyer spruce (*Picea meyeri*) on a sandy substrate in semi-arid grassland, North China, *Trees-Struct. Funct.*, 15, 230–235, doi:[10.1007/s004680100097](https://doi.org/10.1007/s004680100097), 2001.
- 30 Liu, C., Holst, J., Brüggemann, N., Butterbach-Bahl, K., Yao, Z., Han, S., Han, X., and Zheng, X.: Effects of irrigation on nitrous oxide, methane and carbon dioxide fluxes in an Inner Mongolian steppe, *Adv. Atmos. Sci.*, 25, 748–756, doi:[10.1007/s00376-008-0748-3](https://doi.org/10.1007/s00376-008-0748-3), 2008.

- Lyons, T. J. and Halldin, S.: Surface heterogeneity and the spatial variation of fluxes, *Agr. Forest. Meteorol.*, 121, 153–165, doi:[10.1016/j.agrformet.2003.08.031](https://doi.org/10.1016/j.agrformet.2003.08.031), 2004.
- Mahrt, L.: Flux sampling errors for aircraft and towers, *J. Atmos. Oceanic Technol.*, 15, 416–429, doi:[10.1175/1520-0426\(1998\)015<0416:FSEFAA>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<0416:FSEFAA>2.0.CO;2), 1998.
- 5 Mahrt, L.: Surface heterogeneity and vertical structure of the boundary layer, *Bound.-Lay. Meteorol.*, 96, 33–62, doi:[10.1023/a:1002482332477](https://doi.org/10.1023/a:1002482332477), 2000.
- Markkanen, T., Steinfeld, G., Kljun, N., Raasch, S., and Foken, T.: A numerical case study on footprint model performance under inhomogeneous flow conditions, *Meteorol. Z.*, 19, 539–547, doi:[10.1127/0941-2948/2010/0488](https://doi.org/10.1127/0941-2948/2010/0488), 2010.
- 10 Mason, P. J.: The formation of areally-averaged roughness lengths, *Q. J. R. Meteorolog. Soc.*, 114, 399–420, doi:[10.1002/qj.49711448007](https://doi.org/10.1002/qj.49711448007), 1988.
- Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Lee, J. T., Zhao, C., Bakwin, P. S., Munger, J. W., and Hollinger, D. Y.: Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus B*, 58, 344–358, doi:[10.1111/j.1600-0889.2006.00206.x](https://doi.org/10.1111/j.1600-0889.2006.00206.x), 2011.
- 15 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](https://doi.org/10.1029/2006JD008133), 2007.
- Meijninger, W. M. L., Beyrich, F., Lüdi, A., Kohsiek, W., and De Bruin, H. A. R.: Scintillometer-based turbulent fluxes of sensible and latent heat over a heterogeneous land surface – a contribution to LITFASS-2003, *Bound.-Lay. Meteorol.*, 121, 89–110, doi:[10.1007/s10546-005-9022-8](https://doi.org/10.1007/s10546-005-9022-8), 2006.
- Mengelkamp, H. T., Beyrich, F., Heinemann, G., Ament, F., Bange, J., Berger, F., Bosenberg, J., Foken, T., Hennemuth, B., Heret, C., Huneke, S., Johnsen, K. P., Kerschgens, M., Kohsiek, W., Leps, J. P., Liebenthal, C., Lohse, H., Mauder, M., Meijninger, W., Raasch, S., Simmer, C., Spiess, T., 25 Tittebrand, A., Uhlenbrock, J., and Zittel, R.: Evaporation over a heterogeneous land surface – the EVA-GRIPS project, *Bull. Am. Meteorol. Soc.*, 87, 775–786, doi:[10.1175/BAMS-87-6-775](https://doi.org/10.1175/BAMS-87-6-775), 2006.
- Metzger, S., Junkermann, W., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Measuring the 3-D wind vector with a weight-shift microlight aircraft, *Atmos. Meas. Tech.*, 4, 1421–1444, doi:[10.5194/amt-4-1421-2011](https://doi.org/10.5194/amt-4-1421-2011), 2011.
- 30 Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, *Atmos. Meas. Tech.*, 5, 1699–1717, doi:[10.5194/amt-5-1699-2012](https://doi.org/10.5194/amt-5-1699-2012), 2012.
- Meurer, M. and Jiang, Y.: Die Steppen Nordchinas und ihre Belastung durch weide- und land-

wirtschaftliche Landnutzung, *Geographische Rundschau*, 53, 48–52, 2001.

Münel, C. and Roininen, R.: Automatic monitoring of boundary layer structures and depth with ceilometer using a novel robust algorithm, *International Symposium for the Advancement of Boundary Layer Remote Sensing*, Paris, France, 2010.

5 Nagler, P. L., Glenn, E. P., Kim, H., Emmerich, W., Scott, R. L., Huxman, T. E., and Huete, A. R.: Relationship between evapotranspiration and precipitation pulses in a semiarid rangeland estimated by moisture flux towers and MODIS vegetation indices, *J. Arid. Environ.*, 70, 443–462, doi:[10.1016/j.jaridenv.2006.12.026](https://doi.org/10.1016/j.jaridenv.2006.12.026), 2007.

Ogunjemiyo, S. O., Kaharabata, S. K., Schuepp, P. H., MacPherson, I. J., Desjardins, R. L., and
10 Roberts, D. A.: Methods of estimating CO₂, latent heat and sensible heat fluxes from estimates of land cover fractions in the flux footprint, *Agr. Forest. Meteorol.*, 117, 125–144, doi:[10.1016/S0168-1923\(03\)00061-3](https://doi.org/10.1016/S0168-1923(03)00061-3), 2003.

Oke, T. R.: *Boundary layer climates*, Taylor & Francis, London, 435 pp., 1987.

Qi, Y.-C., Dong, Y.-S., Liu, J.-Y., Domroes, M., Geng, Y.-B., Liu, L.-X., Liu, X.-R., and Xiao, H. Y.:
15 Effect of the conversion of grassland to spring wheat field on the CO₂ emission characteristics in Inner Mongolia, China, *Soil Tillage Res.*, 94, 310–320, doi:[10.1016/j.still.2006.08.008](https://doi.org/10.1016/j.still.2006.08.008), 2007.

Quenouille, M. H.: Notes on bias in estimation, *Biometrika*, 43, 353–360, doi:[10.1093/biomet/43.3-4.353](https://doi.org/10.1093/biomet/43.3-4.353), 1956.

R Development Core Team: R: A language and environment for statistical computing, R Foundation for
20 Statistical Computing, Vienna, Austria, 2012.

Raupach, M. R. and Finnigan, J. J.: Scale issues in boundary-layer meteorology: Surface energy balances in heterogeneous terrain, *Hydrol. Processes*, 9, 589–612, doi:[10.1002/hyp.3360090509](https://doi.org/10.1002/hyp.3360090509), 1995.

Ridgeway, G.: *gbm* V1.6-3.2 – Generalized boosted regression models, available under: <http://cran.r-project.org/web/packages/gbm/>, 2012.

25 Ripley, B. D. and Thompson, M.: Regression techniques for the detection of analytical bias, *Analyst*, 112, 377–383, doi:[10.1039/an9871200377](https://doi.org/10.1039/an9871200377), 1987.

Rousseeuw, P. J. and Verboven, S.: Robust estimation in very small samples, *Computation. Stat. Data An.*, 40, 741–758, doi:[10.1016/S0167-9473\(02\)00078-6](https://doi.org/10.1016/S0167-9473(02)00078-6), 2002.

Schaffrath, D., Barthold, F., and Bernhofer, C.: Spatiotemporal variability of grassland vegetation cover in a catchment in Inner Mongolia, China, derived from MODIS data products, *Plant Soil*, 340, 181–198, doi:[10.1007/s11104-010-0465-4](https://doi.org/10.1007/s11104-010-0465-4), 2011.

Schmid, H. P.: Footprint modeling for vegetation atmosphere exchange studies: A review and perspective, *Agr. Forest. Meteorol.*, 113, 159–183, doi:[10.1016/S0168-1923\(02\)00107-7](https://doi.org/10.1016/S0168-1923(02)00107-7), 2002.

Schuepp, P. H., MacPherson, J. I., and Desjardins, R. L.: Adjustment of footprint correction for airborne flux mapping over the FIFE site, *J. Geophys. Res. Atmos.*, 97, 18455–18466, doi:[10.1029/92JD00884](https://doi.org/10.1029/92JD00884), 1992.

5 Shao, C., Chen, J., Li, L., Xu, W., Chen, S., Gwen, T., Xu, J., and Zhang, W.: Spatial variability in soil heat flux at three Inner Mongolia steppe ecosystems, *Agric. For. Meteorol.*, 148, 1433–1443, doi:[10.1016/j.agrformet.2008.04.008](https://doi.org/10.1016/j.agrformet.2008.04.008), 2008.

Sogachev, A. and Lloyd, J.: Using a one-and-a-half order closure model of the atmospheric boundary layer for surface flux footprint estimation, *Bound.-Lay. Meteorol.*, 112, 467–502, 2004.

10 Sorbjan, Z.: Statistics of scalar fields in the atmospheric boundary layer based on large-eddy simulations. Part 2: Forced convection, *Bound.-Lay. Meteorol.*, 119, 57–79, doi:[10.1007/s10546-005-9014-8](https://doi.org/10.1007/s10546-005-9014-8), 2006.

Steffens, M., Kölbl, A., Totsche, K. U., and Kögel-Knabner, I.: Grazing effects on soil chemical and physical properties in a semiarid steppe of Inner Mongolia (China), *Geoderma*, 143, 63–72, doi:[10.1016/j.geoderma.2007.09.004](https://doi.org/10.1016/j.geoderma.2007.09.004), 2008.

15 Strunin, M. A. and Hiyama, T.: Applying wavelet transforms to analyse aircraft-measured turbulence and turbulent fluxes in the atmospheric boundary layer over Eastern Siberia, *Hydrol. Process.*, 18, 3081–3098, doi:[10.1002/hyp.5750](https://doi.org/10.1002/hyp.5750), 2004.

20 Strunin, M. A., Hiyama, T., Asanuma, J., and Ohata, T.: Aircraft observations of the development of thermal internal boundary layers and scaling of the convective boundary layer over non-homogeneous land surfaces, *Bound.-Lay. Meteorol.*, 111, 491–522, 2004.

Stull, R. B.: *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers, Dordrecht, 670 pp., 1988.

Thomas, R. M., Lehmann, K., Nguyen, H., Jackson, D. L., Wolfe, D., and Ramanathan, V.: Measurement of turbulent water vapor fluxes using a lightweight unmanned aerial vehicle system, *Atmos. Meas. Tech.*, 5, 243–257, doi:[10.5194/amt-5-243-2012](https://doi.org/10.5194/amt-5-243-2012), 2012.

25 Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, *Bull. Am. Meteorol. Soc.*, 79, 61–78, doi:[10.1175/1520-0477\(1998\)079<0061:apgtwa>2.0.co;2](https://doi.org/10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2), 1998.

Tukey, J. W.: Bias and confidence in not-quite large samples, *Ann. Math. Stat.*, 29, 614, 1958.

30 Van Kesteren, B., Hartogensis, O. K., van Dinther, D., Moene, A. F., and De Bruin, H. A. R.: Measuring H₂O and CO₂ fluxes at field scales with scintillometry: Part 1 – Introduction and validation of four methods, *Agric. For. Meteorol.*, in press, doi:[10.1016/j.agrformet.2012.09.013](https://doi.org/10.1016/j.agrformet.2012.09.013), 2013.

Vellinga, O. S., Gioli, B., Elbers, J. A., Holtslag, A. A. M., Kabat, P., and Hutjes, R. W. A.: Regional carbon dioxide and energy fluxes from airborne observations using flight-path segmentation based on

- landscape characteristics, *Biogeosciences*, 7, 1307–1321, doi:[10.5194/bg-7-1307-2010](https://doi.org/10.5194/bg-7-1307-2010), 2010.
- Vesala, T., Kljun, N., Rannik, U., Rinne, J., Sogachev, A., Markkanen, T., Sabelfeld, K., Foken, T., and Leclerc, M. Y.: Flux and concentration footprint modelling: State of the art, *Environ. Pollut.*, 152, 653–666, doi:[10.1016/j.envpol.2007.06.070](https://doi.org/10.1016/j.envpol.2007.06.070), 2008.
- 5 Vetter, S., Schaffrath, D., and Bernhofer, C.: Spatial simulation of evapotranspiration of semi-arid Inner Mongolian grassland based on MODIS and eddy covariance data, *Environmental Earth Sciences*, 65, 1567–1574, doi:[10.1007/s12665-011-1187-5](https://doi.org/10.1007/s12665-011-1187-5), 2012.
- Vickers, D. and Mahrt, L.: Quality control and flux sampling problems for tower and aircraft data, *J. Atmos. Oceanic Technol.*, 14, 512–526, doi:[10.1175/1520-0426\(1997\)014<0512:QCAFSP>2.0.CO;2](https://doi.org/10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2),
10 1997.
- Wan, Z. and Li, Z. L.: Radiance-based validation of the V5 MODIS land-surface temperature product, *Int. J. Remote Sens.*, 29, 5373–5395, doi:[10.1080/01431160802036565](https://doi.org/10.1080/01431160802036565), 2008.
- Wang, W. and Rotach, M. W.: Flux footprints over an undulating surface, *Bound.-Layer. Meteorol.*, 136, 325–340, doi:[10.1007/s10546-010-9498-8](https://doi.org/10.1007/s10546-010-9498-8), 2010.
- 15 Webb, E. K., Pearman, G. I., and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, *Q. J. R. Meteorolog. Soc.*, 106, 85–100, doi:[10.1002/qj.49710644707](https://doi.org/10.1002/qj.49710644707), 1980.
- Weil, J., Sullivan, P., Patton, E., and Moeng, C.-H.: Statistical variability of dispersion in the convective boundary layer: Ensembles of simulations and observations, *Bound.-Layer. Meteorol.*, 145, 185–210,
20 doi:[10.1007/s10546-012-9704-y](https://doi.org/10.1007/s10546-012-9704-y), 2012.
- Wiesmeier, M., Barthold, F., Blank, B., and Kögel-Knabner, I.: Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem, *Plant Soil*, 340, 7–24, doi:[10.1007/s11104-010-0425-z](https://doi.org/10.1007/s11104-010-0425-z), 2011.
- Willett, J. B. and Singer, J. D.: Another cautionary note about R^2 : Its use in weighted least-squares regression analysis, *Am. Stat.*, 42, 236–238, 1988.
- Wittmer, M. H. O. M., Auerswald, K., Bai, Y., Schäufele, R., and Schnyder, H.: Changes in the abundance of C3/C4 species of Inner Mongolia grassland: Evidence from isotopic composition of soil and vegetation, *Global Change Biol.*, 16, 605–616, doi:[10.1111/j.1365-2486.2009.02033.x](https://doi.org/10.1111/j.1365-2486.2009.02033.x), 2010.
- Wood, N. and Mason, P.: The influence of static stability on the effective roughness lengths for momentum and heat transfer, *Q. J. R. Meteorolog. Soc.*, 117, 1025–1056, doi:[10.1002/qj.49711750108](https://doi.org/10.1002/qj.49711750108),
25 1991.
- Xiang, G., Huete, A. R., and Didan, K.: Multisensor comparisons and validation of MODIS vegetation indices at the semiarid Jornada experimental range, *IEEE Trans. Geosci. Remote Sens.*, 41, 2368–

- 2381, doi:[10.1109/tgrs.2003.813840](https://doi.org/10.1109/tgrs.2003.813840), 2003.
- Yang, F., Ichii, K., White, M. A., Hashimoto, H., Michaelis, A. R., Votava, P., Zhu, A. X., Huete, A., Running, S. W., and Nemani, R. R.: Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through support vector machine approach, *Remote Sens. Environ.*, 110, 109–122, doi:[10.1016/j.rse.2007.02.016](https://doi.org/10.1016/j.rse.2007.02.016), 2007.
- 5 Ziehn, T., Knorr, W., and Scholze, M.: Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system, *Global Biogeochem. Cy.*, 25, GB2021, doi:[10.1029/2010gb003886](https://doi.org/10.1029/2010gb003886), 2011.

Table 1. Summary of the WSMA flights selected for analysis and related surface conditions. Shown are date, Chinese standard time (CST = Coordinated universal time + 8), flight identifier (ID), length of each flight line l , repetitions rep, cumulated precipitation in a 10 day trailing window P , most frequently occurring land cover classes LC1–LC3, and the enhanced vegetation index EVI immediately below the flight lines. A legend with colour codes for LC and EVI is provided at the bottom. The LC colour code and corresponding abbreviations are identical with Fig. 2.

Date	Time CST	ID	l [km]	rep	P [mm]	LC1	LC2	LC3	EVI
8 July 2009	10:20–10:50	O10	15	3	1.4	54%	35%	6%	35 ± 13%
	12:00–12:50	O12	13	6		32%	30%	21%	35 ± 18%
13 July 2009	11:30–12:10	O8	30	2	5.0	59%	15%	11%	32 ± 10%
	12:40–13:10	O3	21	2		70%	17%	13%	25 ± 6%
15 July 2009	11:30–12:20	O11	11	6	5.0	48%	42%	10%	33 ± 9%
	12:30–13:00	O7	11	4		79%	16%	4%	21 ± 6%
17 July 2009	11:00–11:30	O11	11	4	5.2	54%	40%	4%	36 ± 12%
	12:20–13:00	O7	11	5		82%	10%	5%	21 ± 8%
26 July 2009	12:50–15:30	C1	60	2	13.4	51%	24%	20%	25 ± 7%
	13:10–15:10	C2	63	2		73%	10%	9%	27 ± 9%
30 July 2009	11:00–13:30	C1	60	2	14.3	52%	24%	18%	25 ± 7%
	11:10–13:20	C2	63	2		74%	10%	8%	26 ± 8%

<i>LC</i>	Arable	Bare	Dunes	Irrigated	Marsh	Steppe
<i>EVI</i>	> 20%–25%	> 25%–30%	> 30%–35%	> 35%		

Table 2. Mean length scales \pm SD between repetitions during the WSMA flights selected for analysis. Shown are CBL depth z_i , aerodynamic roughness length z_0 , flight altitude z , thermal blending height z_{TB1} , length scale of surface heterogeneity L_H , Raupach length L_R , the thermal blending lengths L_{TB1} , L_{TB2} , and the upwind distance from the WSMA $L_{80\%}$, where 80 % of the flux contributions are included in the flux footprint.

Date	Time CST	ID	z_i (m)	z_0 (m)	z (m)	z_{TB1} (m)	L_H (m)	L_R (m)	L_{TB1} (m)	L_{TB2} (m)	$L_{80\%}$ (m)
8 Jul 2009	10:20–10:50	O10	1100	0.21 \pm 0.13	59 \pm 4	45 \pm 7	802 \pm 192	1532 \pm 105	1046 \pm 171	255 \pm 5	931 \pm 52
	12:00–12:50	O12	1800	0.07 \pm 0.05	72 \pm 6	24 \pm 2	700 \pm 58	4093 \pm 164	2160 \pm 263	628 \pm 39	1574 \pm 147
13 Jul 2009	11:30–12:10	O8	1900	0.04 \pm 0.05	51 \pm 0	41 \pm 4	1055 \pm 5	4770 \pm 3	1330 \pm 122	999 \pm 12	1440 \pm 127
	12:40–13:10	O3	2100	0.05 \pm 0.07	51 \pm 2	38 \pm 5	858 \pm 95	4292 \pm 236	1134 \pm 45	1108 \pm 220	1305 \pm 64
15 Jul 2009	11:30–12:20	O11	2200	0.06 \pm 0.04	55 \pm 4	11 \pm 2	370 \pm 54	5214 \pm 508	1950 \pm 223	1368 \pm 304	1470 \pm 168
	12:30–13:00	O7	2100	0.26 \pm 0.19	57 \pm 7	17 \pm 6	298 \pm 76	3507 \pm 82	984 \pm 117	1207 \pm 132	1081 \pm 128
17 Jul 2009	11:00–11:30	O11	1400	1.13 \pm 0.81	48 \pm 1	22 \pm 0	366 \pm 37	1589 \pm 65	788 \pm 74	440 \pm 28	720 \pm 74
	12:20–13:00	O7	1400	0.05 \pm 0.06	52 \pm 2	19 \pm 8	507 \pm 219	3136 \pm 169	1381 \pm 54	950 \pm 159	1296 \pm 150
26 Jul 2009	12:50–15:30	C1	2500	1.75 \pm 1.91	97 \pm 4	85 \pm 68	1458 \pm 913	2626 \pm 284	1974 \pm 562	752 \pm 196	991 \pm 331
	13:10–15:10	C2	2500	2.30 \pm 2.12	102 \pm 5	83 \pm 51	1459 \pm 632	2430 \pm 382	1921 \pm 325	916 \pm 310	945 \pm 191
30 Jul 2009	11:00–13:30	C1	1600	0.48 \pm 0.34	56 \pm 3	46 \pm 21	1653 \pm 161	3097 \pm 340	2403 \pm 1438	1015 \pm 22	1042 \pm 100
	11:10–13:20	C2	1600	0.11 \pm 0.01	54 \pm 0	51 \pm 13	2615 \pm 84	3798 \pm 735	2853 \pm 601	1852 \pm 51	1206 \pm 0

Table 3. Median land cover specific flux estimates of H and LE from the LTFM procedure over all flight patterns \pm median spatial variability within the respective land cover. Also shown are the corresponding median ensemble random uncertainties $\sigma_{\text{ens}}(H)$, $\sigma_{\text{ens}}(\text{LE})$ and land cover specific sample size N .

Land cover	H (W m^{-2})	LE (W m^{-2})	$\sigma_{\text{ens}}(H)$	$\sigma_{\text{ens}}(\text{LE})$	N
Bare soil	193 ± 32	136 ± 38	1 %	1 %	22049
Sand dunes	188 ± 37	144 ± 55	1 %	1 %	43424
Marshland	125 ± 55	230 ± 59	1 %	1 %	20722
Steppe	202 ± 40	138 ± 46	< 1 %	< 1 %	321956
Mountain meadow	114 ± 47	260 ± 69	1 %	1 %	25175
Settlements	172 ± 40	155 ± 46	3 %	5 %	1404
Rainfed agriculture	183 ± 35	147 ± 40	1 %	1 %	17024
Irrigated agriculture	116 ± 32	224 ± 41	5 %	5 %	1068

Table 4. Median systematic- and random uncertainty terms (in parenthesis) for a single flux observation or grid cell throughout the LTFM procedure.

Source	<i>H</i>	LE
Instrumentation and hardware	0 % (8 %)	0 % (7 %)
Turbulence sampling	0 % (57 %)	0 % (121 %)
Spatio-temporal analysis	2 % (40 %)	4 % (47 %)
BRT residuals	0 % (5 %)	0 % (6 %)
BRT response function	11 % (69 %)	18 % (77 %)
BRT state variables	13 % (77 %)	14 % (75 %)

Table 5. Mean meteorological conditions \pm SD between repetitions during the WSMA flights selected for analysis. Shown are cloud cover CC, shortwave down-welling radiation S_{\downarrow} , mixing ratio MR, horizontal wind speed u , wind direction DIR, virtual potential temperature θ_v , surface temperature T_s , and the SD of the surface temperature σ_{T_s} .

Date	Time CST	ID	CC	S_{\downarrow} (Wm^{-2})	MR (gkg^{-1})	u (ms^{-1})	DIR ($^{\circ}$)	θ_v (K)	T_s ($^{\circ}\text{C}$)	σ_{T_s} (K)
8 Jul 2009	10:20–10:50	O10	7/8	842 ± 55	6.4 ± 0.0	3.2 ± 0.3	221 ± 12	312.0 ± 0.4	40.8 ± 1.2	8.1 ± 0.4
	12:00–12:50	O12	7/8	773 ± 53	5.9 ± 0.2	6.5 ± 0.3	320 ± 2	313.7 ± 0.3	39.8 ± 1.0	8.8 ± 0.4
13 Jul 2009	11:30–12:10	O8	4/8	810 ± 16	9.3 ± 0.1	8.3 ± 0.4	291 ± 6	309.7 ± 0.1	41.9 ± 0.3	6.4 ± 0.1
	12:40–13:10	O3	4/8	838 ± 6	8.6 ± 0.1	6.9 ± 0.6	297 ± 6	311.2 ± 0.3	45.8 ± 0.2	5.3 ± 0.8
15 Jul 2009	11:30–12:20	O11	7/8	796 ± 72	7.1 ± 0.1	7.0 ± 0.8	253 ± 9	315.3 ± 0.2	42.1 ± 1.1	5.3 ± 0.7
	12:30–13:00	O7	7/8	843 ± 56	6.8 ± 0.0	5.8 ± 0.2	255 ± 8	316.7 ± 0.2	50.8 ± 0.7	4.0 ± 0.5
17 Jul 2009	11:00–11:30	O11	7/8	589 ± 39	9.4 ± 0.1	2.7 ± 0.2	102 ± 5	309.3 ± 0.2	35.9 ± 0.5	4.8 ± 0.3
	12:20–13:00	O7	7/8	682 ± 122	11.2 ± 0.2	5.9 ± 0.4	144 ± 4	310.4 ± 0.1	40.3 ± 2.5	4.5 ± 0.9
26 Jul 2009	12:50–15:30	C1	7/8	668 ± 46	9.6 ± 0.3	2.9 ± 0.1	174 ± 5	312.8 ± 0.5	36.4 ± 1.2	4.8 ± 0.7
	13:10–15:10	C2	7/8	747 ± 67	9.1 ± 0.1	2.7 ± 0.3	178 ± 23	313.0 ± 0.4	36.5 ± 0.5	3.7 ± 0.7
30 Jul 2009	11:00–13:30	C1	7/8	715 ± 82	11.6 ± 0.3	4.3 ± 0.7	159 ± 15	311.8 ± 0.9	34.6 ± 5.1	4.1 ± 0.5
	11:10–13:20	C2	7/8	567 ± 11	11.6 ± 0.0	4.9 ± 0.9	154 ± 9	311.3 ± 1.2	32.8 ± 1.0	2.7 ± 0.5

Table 6. Mean turbulence statistics \pm SD between repetitions during the WSMA flights selected for analysis: friction velocity u_* , sensible heat flux H , and latent heat flux LE, Monin-Obukhov length L_v , SD of vertical wind σ_w , and convective velocity w_* .

Date	Time CST	ID	u_* (ms ⁻¹)	H (Wm ⁻²)	LE (Wm ⁻²)	L_v (m)	σ_w (ms ⁻¹)	w_* (ms ⁻¹)
8 Jul 2009	10:20–10:50	O10	0.31 \pm 0.03	154 \pm 20	194 \pm 45	-14 \pm 3	0.88 \pm 0.03	1.82 \pm 0.08
	12:00–12:50	O12	0.46 \pm 0.07	199 \pm 14	110 \pm 36	-38 \pm 16	0.97 \pm 0.04	2.29 \pm 0.05
13 Jul 2009	11:30–12:10	O8	0.52 \pm 0.09	290 \pm 47	196 \pm 43	-40 \pm 24	0.99 \pm 0.05	2.65 \pm 0.13
	12:40–13:10	O3	0.42 \pm 0.12	288 \pm 18	86 \pm 51	-23 \pm 19	0.95 \pm 0.01	2.70 \pm 0.07
15 Jul 2009	11:30–12:20	O11	0.47 \pm 0.07	176 \pm 17	138 \pm 30	-46 \pm 23	0.83 \pm 0.05	2.36 \pm 0.07
	12:30–13:00	O7	0.52 \pm 0.12	310 \pm 28	65 \pm 12	-38 \pm 19	1.07 \pm 0.11	2.77 \pm 0.08
17 Jul 2009	11:00–11:30	O11	0.39 \pm 0.07	156 \pm 26	46 \pm 67	-30 \pm 15	0.79 \pm 0.04	1.93 \pm 0.09
	12:20–13:00	O7	0.41 \pm 0.07	206 \pm 18	48 \pm 45	-27 \pm 13	0.83 \pm 0.06	2.11 \pm 0.07
26 Jul 2009	12:50–15:30	C1	0.33 \pm 0.09	120 \pm 33	300 \pm 76	-20 \pm 10	0.94 \pm 0.11	2.25 \pm 0.21
	13:10–15:10	C2	0.37 \pm 0.05	117 \pm 11	227 \pm 2	-29 \pm 8	0.94 \pm 0.04	2.22 \pm 0.06
30 Jul 2009	11:00–13:30	C1	0.43 \pm 0.13	107 \pm 76	194 \pm 27	-57 \pm 14	0.75 \pm 0.12	1.78 \pm 0.41
	11:10–13:20	C2	0.38 \pm 0.05	71 \pm 4	223 \pm 11	-50 \pm 19	0.68 \pm 0.04	1.65 \pm 0.02



Fig. 1. Location of the Xilin River Catchment in the Inner Mongolia Autonomous Region, China (modified after Steffens et al., 2008).

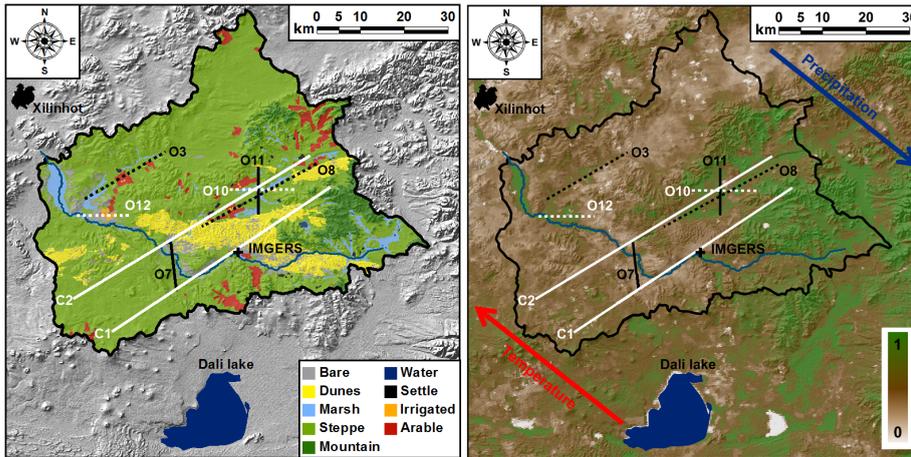


Fig. 2. Maps of the Xilin River Catchment (black boundary), with the IMGERS research station and pairs of flight lines. Left: land cover classification (modified after Wiesmeier et al., 2011) over a digital elevation model (Shuttle Radar Topography Mission, Tile 60_04, data version 4.1, Jarvis et al., 2008). The colour codes are abbreviated for bare soil (Bare), marshland (Marsh), generic steppe (Steppe), mountain meadow (Mountain), settlements (Settle), irrigated agriculture (Irrigated), and rainfed agriculture (Arable). Right: MODIS enhanced vegetation index of 20 July 2009 with a colour bar ranging from $0 < EVI < 1$.

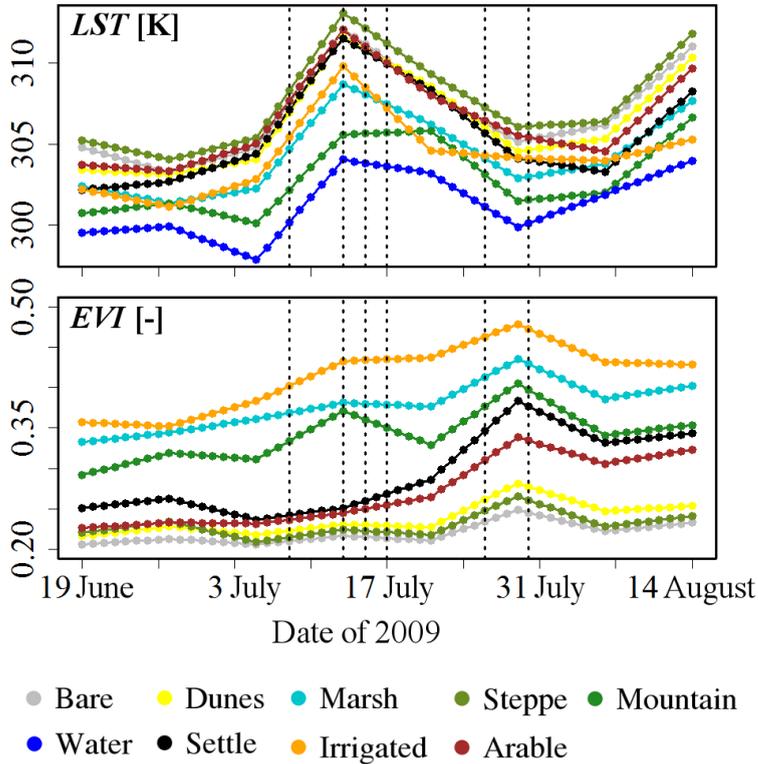


Fig. 3. Change of land surface temperature (top) and enhanced vegetation index (bottom) for each land cover class throughout the study period. The land cover colour code and corresponding abbreviations are identical with Fig. 2. The land cover “Water” is not present for the EVI, because water absorbs strongly in the near infrared, leading to negative EVI values that are not indicative of vegetation greenness.

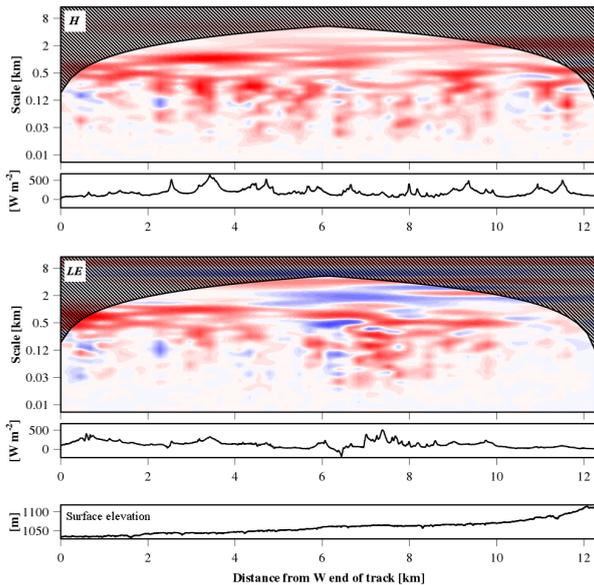


Fig. 4. Wavelet cross-scalograms for the sensible heat flux (top panel) and the latent heat flux (centre panel) along flight pattern O12 on 8 July 2009, 12:16–12:24 CST. The colour palette changes from blue (downward fluxes) over white (neutral) to red (upward fluxes). The shaded areas identify the cone of influence. Beneath the cross-scalograms the integrated flux over all scales is shown for each overflown 90 m cell of the land cover grid. The surface elevation along the flight pattern is displayed in the bottom panel.

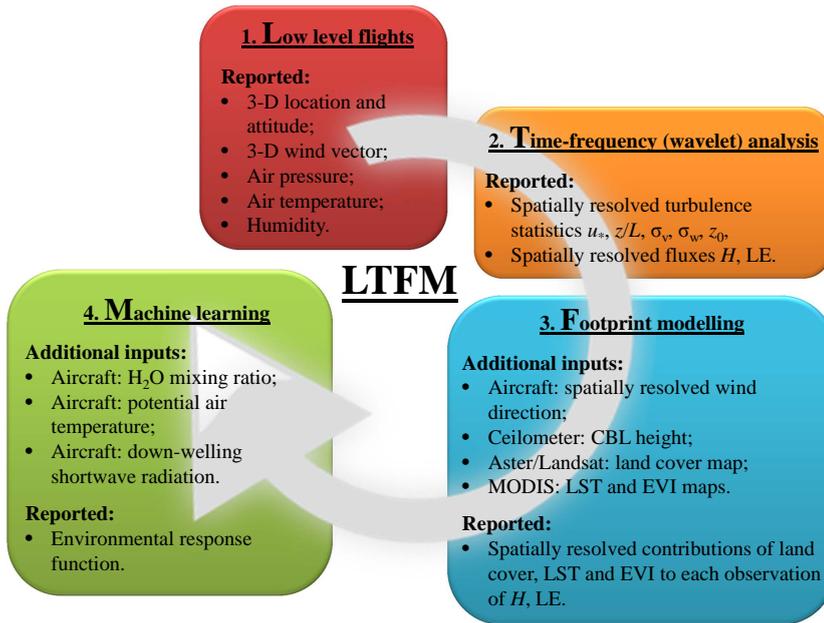


Fig. 5. Flow chart showing how input and reported data streams are processed along the four principal steps of the LTFM method. Additional detail is provided in Sects. 2.4.4 and 4, and a summary of all notation can be found in Appendix A.

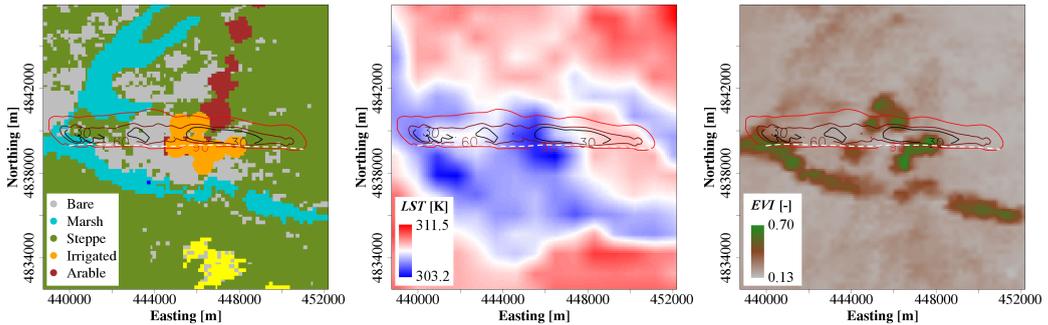


Fig. 6. Flight along pattern O12 on 8 July 2009, 12:16–12:24 CST (white dashed line). The composite flux footprint along the flight line (30 %, 60 %, 90 % contour lines) is superimposed over maps of land cover (left panel), land surface temperature (LST, center panel), and enhanced vegetation index (EVI, right panel). The land cover colour code and corresponding abbreviations are identical with Fig. 2.

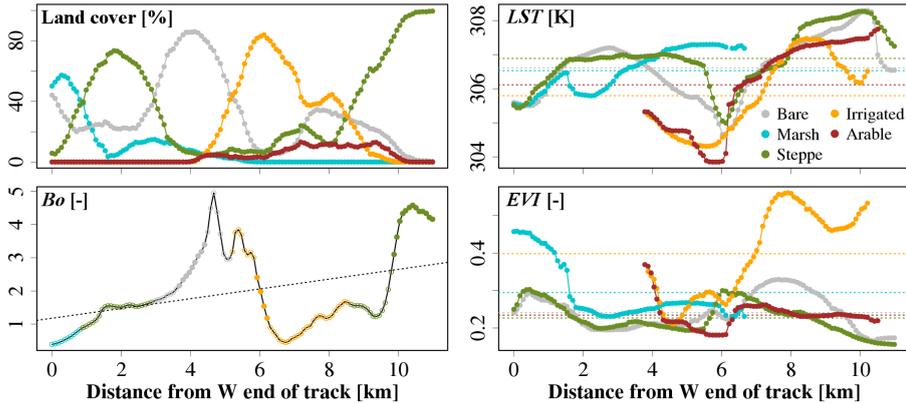


Fig. 7. Biophysical surface properties in the footprint of each observation ($N = 124$) along the flight pattern O12 on 8 July 2009, 12:16–12:24 CST, summarized by land cover. Shown are (clockwise from top right panel) land surface temperature, enhanced vegetation index, Bowen ratio, and the land cover fraction in the footprint. The dashed lines are land cover averages for LST and EVI, and the spatial trend for Bo. The land cover colour code and corresponding abbreviations are identical with Fig. 2.

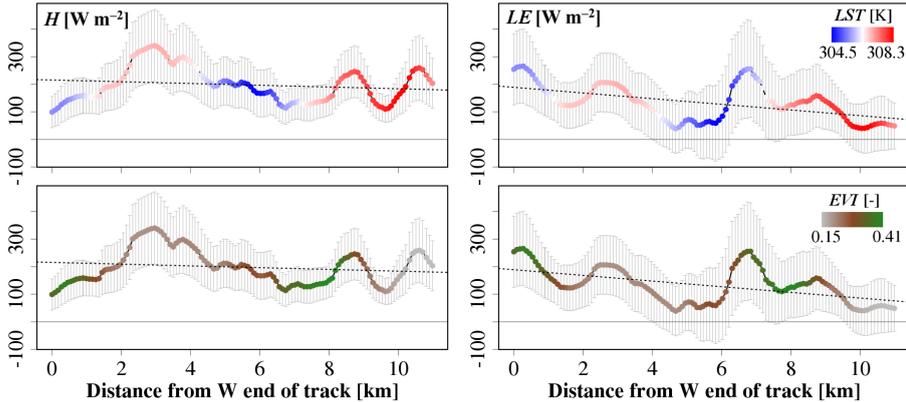


Fig. 8. Sensible heat flux (left panels) and latent heat flux (right panels) along the flight pattern O12 on 8 July 2009, 12:16–12:24 CST. Also shown is the random sampling error (error bars) for each observation ($N = 124$), and the spatial trend (dashed line). The top and bottom panels show the land surface temperature and the enhanced vegetation index in the footprint of each observation, respectively.

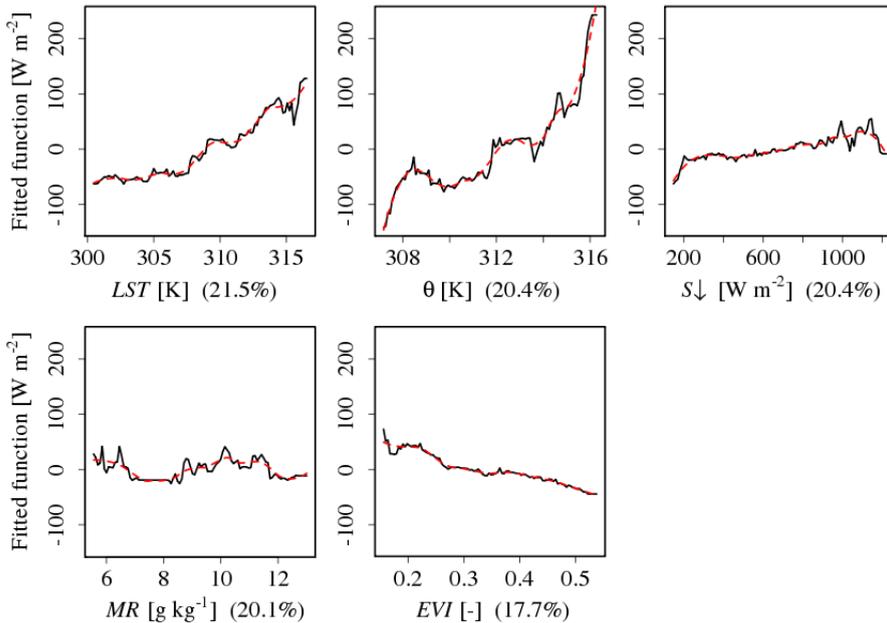


Fig. 9. Boosted regression tree partial response plots of H for all five state variables in order of their relative importance (in braces). The fitted function (black) shows the variable response of the BRT over the range of one individual state variable, while the remaining state variables are held at an average, constant value. The red dashed line is a smoothed representation of the fitted function (locally weighted polynomial regression).

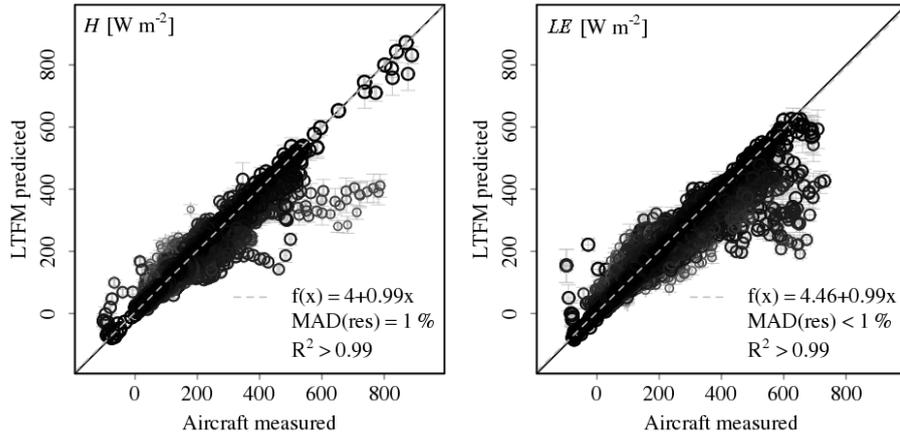


Fig. 10. Maximum likelihood functional relationships between $N = 8446$ aircraft observation and LTFM predictions of sensible heat flux (left) and latent heat flux (right). **The weight of each data point in the relationship is represented by the size of the circles.** The error bars show the cross-validation residuals for the LTFM predictions, and the ensemble random sampling error for the aircraft measurement. The 99.9% confidence intervals are too narrow to be displayed properly.

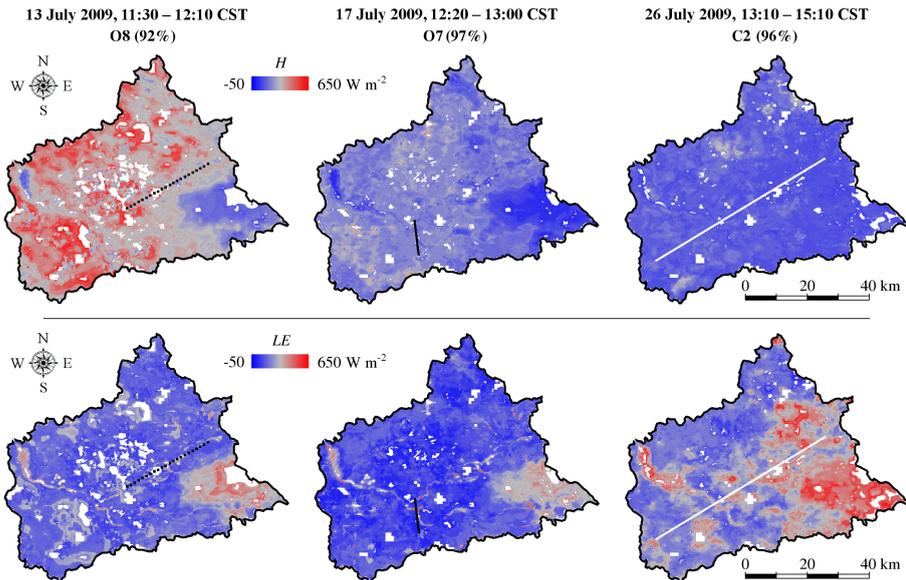


Fig. 11. Maps of LTFM predicted fluxes of sensible heat (H , top) and latent heat (LE , bottom) on 13, 17 and 26 July 2009 (left to right). The colour gradient from blue over grey to red represents values that are lower, equal to, or greater than the average of the values, respectively (see legend). Percentages in braces after the flight ID indicate the spatial coverage of the prediction throughout the catchment. Meteorological state variables from the superimposed flight lines are used in the respective LTFM prediction (illustration identical with Fig. 2).

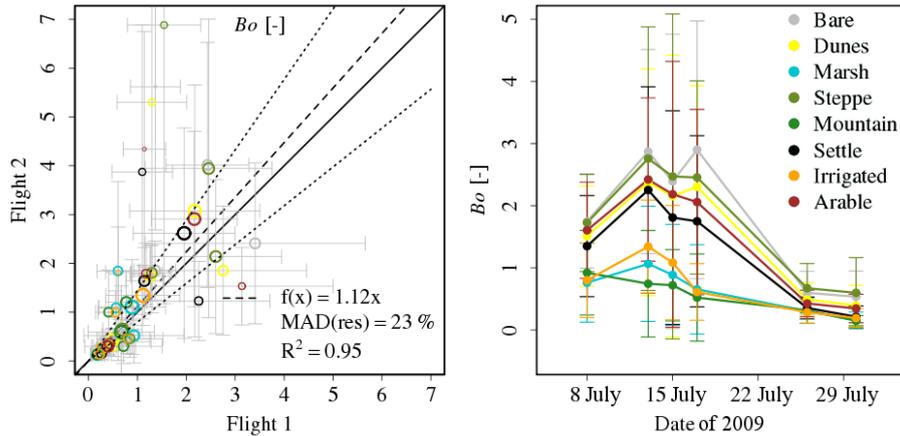


Fig. 12. **Left:** MLFR of Bowen ratio between the first and the second flight pattern on every measurement day. **The weight of each data point in the relationship is represented by the size of the circles.** **Right:** time series of Bo for different land covers throughout the measurement campaign. **In both images** the error bars represent the Gaussian sum of the natural variability in each land cover class and the ensemble random error in the LTFM procedure. The land cover colour code and corresponding abbreviations are identical with Fig. 2.

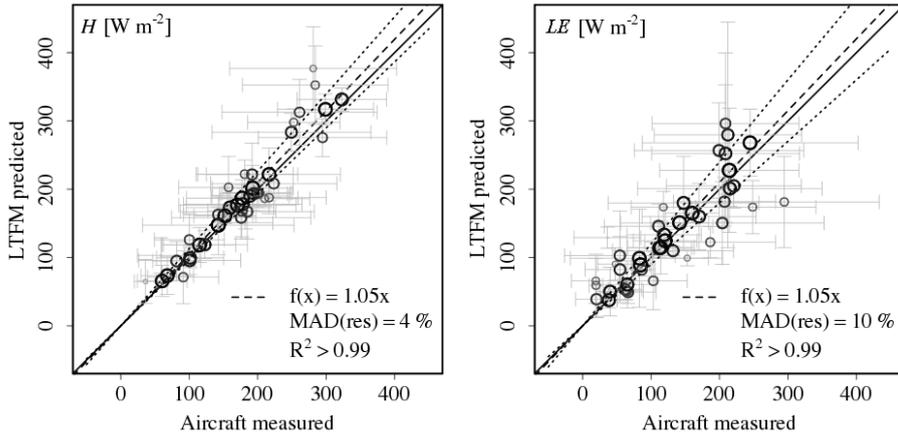


Fig. 13. MLFRs of median observed and predicted fluxes along 42 flight lines. The error bars correspond to the variability of the fluxes along the flight line, and the weight of each data point in the relationship is represented by the size of the circles.