**Hydrology and
Earth System
Sciences
Discussions**

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

A framework to utilize turbulent flux measurements for mesoscale models and remote sensing applications

W. Babel[1], S. Huneke[2], and T. Foken[1]

[1]Department of Micrometeorology, University of Bayreuth, 95440 Bayreuth, Germany
[2]Anemos GmbH, Bunsenstraße 8, 21365 Adendorf, Germany

Correspondence to: W. Babel (wolfgang.babel@uni-bayreuth.de)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Abstract

Meteorologically measured fluxes of energy and matter between the surface and the atmosphere originate from a source area of certain extent, located in the upwind sector of the device. The spatial representativeness of such measurements is strongly influenced by the heterogeneity of the landscape. The footprint concept is capable of linking observed data with spatial heterogeneity. This study aims at upscaling eddy covariance derived fluxes to a grid size of 1 km edge length, which is typical for mesoscale models or low resolution remote sensing data.

Here an upscaling strategy is presented, utilizing footprint modelling and SVAT modelling as well as observations from a target land-use area. The general idea of this scheme is to model fluxes from adjacent land-use types and combine them with the measured flux data to yield a grid representative flux according to the land-use distribution within the grid cell. The performance of the upscaling routine is evaluated with real datasets, which are considered to be land-use specific fluxes in a grid cell. The measurements above rye and maize fields stem from the LITFASS experiment 2003 in Lindenberg, Germany and the respective modelled timeseries were derived by the SVAT model SEWAB. Contributions from each land-use type to the observations are estimated using a forward lagrangian stochastic model. A representation error is defined as the error in flux estimates made when accepting the measurements unchanged as grid representative flux and ignoring flux contributions from other land-use types within the respective grid cell.

Results show that this representation error can be reduced up to 56 % when applying the spatial integration. This shows the potential for further application of this strategy, although the absolute differences between flux observations from rye and maize were so small, that the spatial integration would be rejected in a real situation. Corresponding thresholds for this decision have been estimated as a minimum mean absolute deviation in modelled timeseries of the different land-use types with $35\,\mathrm{W\,m^{-2}}$ for the sensible heat flux and $50\,\mathrm{W\,m^{-2}}$ for the latent heat flux. Finally, a quality flagging

scheme to classify the data with respect to representativeness for a given grid cell is proposed, based on an overall flux error estimate. This enables the data user to infer the uncertainty of mesoscale models and remote sensing products with respect to ground observations. Major uncertainty sources remaining are the lack of an adequate method for energy balance closure correction as well as model structure and parameter estimation, when applying the model for surfaces without flux measurements.

1 Introduction

Long-term modelling of ecosystem fluxes between the terrestrial surface with its particular vegetation and the atmosphere are now widely used facilities of the scientific communities. The need for that arises not only from the desire for better understanding of ecosystem processes, but also from society's request for water and greenhouse gas budgets. Within these data networks the eddy covariance (EC) method stands out as the method of choice, utilized within large flux networks with the mentioned objectives like FLUXNET (Baldocchi et al., 2001). A typical feature of such projects is the spatial integration of flux data in order to derive budgets for entire ecosystem types (e.g. Jung et al., 2009). While observations are too scarce for regional estimates, integration can be achieved via remote sensing data and mesoscale modelling using flux data as the ground truth. These applications work on grid sizes of at least 1 km edge length. Therefore the point measurements have to be related to a certain grid cell, reflecting its properties. The representativeness of the data depends on the heterogeneity within this area, but also on the heterogeneity within the footprint of the measurements.

As upscaling is a problem occurring almost everywhere in earth science, there are lots of approaches to the issue of proper aggregation. The nonlinearity between processes and driving variables, as well as spatial heterogeneity, pose a major challenge for these approaches (Chen et al., 2009). While simple parameter aggregation schemes via mean values or effective parameters (using the "blending height" concept after Mason, 1988) are widely used, there are more comprehensive flux aggregation

5167

methods available (e.g. Hasager and Jensen, 1999). Promising approaches stem from coupling of models, given that coupling occurs via fluxes (Best et al., 2004), and mosaic or subgrid approaches (Avissar and Pielke, 1989; Mölders et al., 1996) should be favoured over parameter averaging wherever possible. A sound basis for estimation of representativeness error, related to both measurement footprints and grid size, is the sensor location bias given by Schmid (1997); Schmid and Lloyd (1999). Further methods were proposed by Tolk et al. (2008), who calculate a representation error of a course grid cell as the standard deviation of carbon dioxide fluxes from the respective finer grid cells. They aim at scales beginning from grid cells with an edge length of 2 km up to 100 km. Only a few recent papers combine the scale of EC measurements and remote sensing data or mesoscale models in a physical way, i.e. by utilization of footprint or SVAT modelling. Some studies develop parameterization schemes for regional fluxes from observations aided by high resolution remote sensing data (e.g. Su, 2002; Ma et al., 2006). Recently, Chen et al. (2009) have offered a scaling methodology based on the footprint climatology of EC field sites and high resolution remote sensing data. They found better agreement between remotely sensed and EC derived gross primary production (GPP) by weighting the former with the footprint climatology than by equal weighting of the grid cells. This approach was also applied to a Chinese field site (Chen et al., 2010).

The footprint approach, which has recently been widely accepted, relates measured data to its sources. It is expected to play a crucial role for matching the scales of EC measurements and remote sensing data or mesoscale models (e.g. Schmid, 1997). Schmid (2002) and Vesala et al. (2008) offer sound overviews of the existing concepts. Validation concepts exist as well (Foken and Leclerc, 2004), but this issue has yet to be investigated (Vesala et al., 2008). Comparisons in Göckede et al. (2005) and Kljun et al. (2002) show better performance of Lagrangian stochastic models over analytical models. Further comparisons of Lagrangian stochastic models with LES (Large Eddy Simulation) model predictions on a basis of 2-D footprint functions showed good agreement for intermediate measurement heights and convective

5168

conditions (Markkanen et al., 2009). Also Finn et al. (1996) and Hsieh and Katul (2009) assign reasonable performance to the Lagrangian stochastic forward model approach on step change heterogeneity, even if this model type is only valid for horizontal homogeneous flow conditions due to the inverted plume assumption. However, not all flow characteristics can be tackled by the footprint concept, as Foken and Leclerc (2004) state that influences of remote obstacles on measurements have been found, even if the disturbed region is beyond the predicted source area. But nevertheless, footprint analysis was combined with measurement quality in order to characterise complex study sites (Göckede et al., 2004, 2006). This scheme was applied successfully for CarboEurope sites (Rebmann et al., 2005; Göckede et al., 2008) and on the Tibetan Plateau (Metzger et al., 2006). Similar procedures were also conducted for FLUXNET sites by Chen et al. (2009), but the scope was more quantitative and hints at the upscaling issue: monthly and annual uncertainties in EC fluxes from a 59-year-old Douglas-fir stand were attributed to variations in footprint climatology and estimated to be approximately 15–20 %. The authors state that footprint-weighted EC fluxes can be used to estimate the bias between spatially-explicit ecological models and tower-based remote sensing at finer scales.

The aim of this study is to combine footprint analysis and SVAT modelling to enhance representativeness of EC-measured turbulent fluxes on a grid level of 1 km edge length, which is a typical size for mesoscale models assisted by moderate resolution remote sensing data. Such models are utilized for comprehensive assessment of energy and water budgets on a regional scale. This approach is useful for field sites with high sensor location biases, where differences in fluxes from various patches are expected to exceed model uncertainties. The need for such algorithms is underpinned by e.g. Avissar (1995); Raupach and Finnigan (1995); Baldocchi et al. (2005); Kim et al. (2006). In this paper, a new upscaling strategy is proposed and evaluated in a case study. The data and models used for this case study are described in Sect. 2. Furthermore, specific measures for representativeness of flux measurements are defined and the new upscaling scheme is explained in Sect. 3. Section 4 starts with an evaluation approach

for this case study and its results are presented including considerations about uncertainty. Finally the implications are summarized in Sect. 5.

2 Methods

2.1 Site description

The investigated datasets were gathered within the LITFASS-2003 experiment from 19 May to 17 June 2003, in a rural landscape around the Meteorological Observatory Lindenberg of the German Meteorological Service (DWD) and its boundary-layer field site Falkenberg, 52°10′01″ N and 14°07′27″ E (Beyrich and Adam, 2007). Fourteen micrometeorological stations were operated over different surfaces, mainly agricultural crops and forests, aiming at evapotranspiration estimates over heterogeneous landscapes (Beyrich and Mengelkamp, 2006; Mengelkamp et al., 2006). Weather conditions were fairly dry during LITFASS-2003, interrupted only by scarce rainfall episodes, which led to a huge variability of water availability and evapotranspiration (Beyrich and Mengelkamp, 2006).

For this investigation, datasets were used from three adjacent farmland sites (from North to South: A4, A5 and A6). While there was no obstacle between A4 and A5, A5 and A6 were separated by a track with a hedgerow. Serving as cropland, A5 was cultivated with rye, and A4 and A6 with maize, during the LITFASS-2003 campaign. All sites were equipped as a full energy balance station, the turbulent fluxes were measured using a Campbell CSAT3 sonic anemometer and a Campbell KH20 hygrometer for A4, a Metek USA-1 sonic anemometer together with a KH20 for A5 and a CSAT3 in combination with a LI-COR LI-7500 CO_2/H_2O gas analyser for A6. Further measurements include radiation with all components for A5 (Kipp & Zonen CNR1 net radiometer) and A6 (Eppley PIR Pyrgeometer, Kipp & Zonen CM24 albedometer), while for A4 only downwelling short-wave radiation (Campbell SP1110) and net radiation (REBS Q7) were recorded. Soil heat flux as well as soil temperature and moisture profiles

were obtained by various instruments (Rimco CN3, Hukseflux HFP, REBS HFT, Campbell 107 Probe, Pt100 and IMKO Trime EZ). Additional site-specific plant physiological and physical parameters such as leaf area index (LAI), canopy height and soil texture were recorded during the campaign, for details see Mauder et al. (2006).

As the campaign was conducted during a growing phase of the respective cereals, canopy height and therefore displacement height and roughness length varied throughout the experiment. For A5 (rye), canopy height began with 90 cm, peaking with 150 cm on 2 June with a final height of 130 cm, recorded on 12 June. A6 (maize) exhibits a much larger variability, starting with 9 cm which could roughly be attributed to bare soil conditions, and growing continuously to 60 cm on 10 June as the last record. Nearly the same pattern shows A4 (maize) with canopy heights ranging from 5 cm to 75 cm.

2.2 Postprocessing of turbulent flux data

The dataset of half-hour turbulent fluxes ($n = 1392$) was calculated using the software package TK2 (Mauder and Foken, 2004), and the processing of LITFASS-2003 data is described in detail in Mauder et al. (2006). Turbulent fluxes were filtered, excluding fluxes with poor quality (data with quality flags of 7–9 were excluded using a scheme ranging from 1 to 9 by Foken et al., 2004). Fetch analysis by Mauder et al. (2006) revealed homgeneous flow conditions and no internal boundary layers for A5 and negligible influence on A4 within the measurement heights $z_m = 2.9$ m and $z_m = 3.25$ m, respectively. The A6 (maize) turbulence complex, however, was situated near the track accompanied by a hedgerow, leading to a disturbed flow field for a wind sector from WNW to ENE. Consequently, A6 data was filtered for the disturbed sector. Furthermore, the contribution of the target land-use was determined by footprint analysis (see Sect. 2.3), the respective half-hour flux values were excluded depending on wind direction and stratification and using a threshold of 80 % for the target contribution, which poses a representative measurement after Göckede et al. (2008).

The energy balance closure (EBC) for LITFASS-2003 sites was investigated by Mauder et al. (2006), and Foken et al. (2010) focused on this issue. The local closure

of the three sites used here ranged from 60 % to 70 % of the available energy, while the SVAT (Soil – Vegetation – Atmosphere – Transport) scheme used in this study closes the energy balance by definition (see Sect. 2.4). This fact compromises direct comparisons of turbulent flux observations and modelled fluxes, as already shown for these sites by Kracher et al. (2009). Therefore the sensible and the latent heat flux are corrected following Twine et al. (2000) as suggested in Foken (2008): the residual *Res* of the energy balance is distributed among the turbulent fluxes while the Bowen ratio *Bo* was preserved yielding the energy balance corrected heat fluxes $Q_{H,EBC}$ and $Q_{E,EBC}$. In order to avoid unreasonable huge corrections and artificial spikes, the correction was not applied when *Bo* was negative or at least one of the turbulent fluxes fail to exceed the measurement accuracy, which was assumed to be $10 \, \text{W m}^{-2}$ in absolute values. Instead, such values were excluded from further analysis.

After all filtering and energy balance correction, missing values sum up to 58–63 % for $Q_{H,EBC}$ and $Q_{E,EBC}$, and an even higher missing fraction in the case of A6 due to the wind sector filtering. Most of these gaps, however, occur during the night, where from experience the fluxes are known to be low. Although this fact is somewhat comforting, it raises the need for gapfilling to obtain unbiased results of mean fluxes on a daily basis or longer time scales. Therefore the energy balance corrected turbulent fluxes are gapfilled with calibrated SVAT model runs for further analysis. The calibration procedure is described in Sect. 2.4 and the influence of gapfilling on the results is discussed in Sect. 4.4.

2.3 Footprint analysis

In this study a Lagrangian stochastic forward model is used to estimate two-dimensional contributions of source areas. While a general description of the model is given by Rannik et al. (2003), it is used with a simplified parameterisation of turbulent flow according to conditions for low vegetation as used in Göckede et al. (2005). As Lagrangian stochastic models require high computational costs, source weight functions

for each half-hour measurement were picked from precalculated tables following a procedure as used in Göckede et al. (2004, 2008).

As a forward model depends on the inverted plume assumption, horizontal homogeneity is required in principle. Experiences from past investigations and validations, however, affirm its applicability even under heterogeneous conditions (e.g. Rebmann et al., 2005; Göckede et al., 2005, 2008; Markkanen et al., 2009, 2010). Further limits stem from flow distortion due to large obstacles like hedges, creating flow patterns and internal boundary layers which cannot be resolved by this model.

2.4 SVAT modelling

As the proposed upscaling scheme needs additional modelling, a SVAT scheme called SEWAB, developed by Mengelkamp et al. (1999) in the former GKSS Research Center, Geesthacht, Germany, is implemented. This model has been chosen due to its energy balance closure technique, which was well rated in a comparison study of Kracher et al. (2009). Here only the principles of the model structure will be highlighted in the following, for further information see the publications mentioned above as well as Kracher et al. (2009). The application and derivation of model parameters is described in Sect. 2.4.2.

2.4.1 SEWAB model structure

Momentum flux (u_*) and sensible heat flux (Q_H) were calculated using a bulk approach

$$u_* = \sqrt{C_D} \cdot u(z) \qquad (u(0) = 0) \tag{1}$$

$$Q_H = \rho \cdot c_p \cdot C_H \cdot u(z) \cdot (T_g - T(z)) \tag{2}$$

with the drag coefficient C_D, the Stanton number C_H and c_p as the specific heat capacity. $u(z)$ and $T(z)$ denote wind velocity and air temperature at height z, while T_g represents the surface temperature. The Stanton number C_H is parameterised depending

on aerodynamic and thermal roughness lengths as well as atmospheric stratification following Louis (1979). For the latent heat flux Q_E a bulk approach also applies in analogy to Eq. (2) in general, but the flux is split up into evaporation from bare soil and vegetated surface flux, the latter again composed of evaporation from wet foliage and transpiration from dry foliage. While for evaporation from bare soil and wet foliage the Dalton number equals the Stanton number C_H, the transpiration drag coefficient was parameterised with the aerodynamic resistance and the stomatal resistance after Noilhan and Planton (1989). Similar, the soil heat flux Q_G is estimated by

$$Q_G = \lambda \left(\frac{T_g - T_{S1}}{\Delta z_{S1}} \right) \tag{3}$$

with the thermal conductivity λ and the first soil layer temperature T_{S1} with a thickness of $\Delta z_{S1} = 2\,\text{cm}$. Soil temperature distribution is described by the diffusion equation and vertical soil water movement is governed by the Richards equation. Relationships between soil moisture characteristics were used from Clapp and Hornberger (1978). The net radiation is written as

$$Q_R^* = -R_{swd}(1 - a) - R_{lwd} + \varepsilon \sigma T_g^4 \tag{4}$$

with the albedo a, the emissivity ε, the Stefan-Boltzmann constant σ and the net radiation Q_R^*. Downwelling short-wave and long-wave radiation (R_{swd} and R_{lwd}) were not parameterised for this dataset but prescribed as a part of the forcing data. Finally, the energy balance is given by

$$-Q_R^* = Q_H + Q_E + Q_G \tag{5}$$

which is closed in SEWAB by definition. While all components are given with separate equations without using the balance equation, the closure is achieved by an iteration of the surface temperature T_g until the residual disappears. Thus, instead of charging one flux to serve as balance residual, the discrepancy is shared by all fluxes sensitive to T_g. Directly affected are Q_H, Q_G, and the upwelling long-wave radiation (see Eqs. (2), (3) and (4)), but also Q_E via the temperature dependent specific humidity of saturation.

2.4.2 SEWAB parameter estimation

The model runs were achieved with two different sets of parameter for each station, an optimised and a "realistic" set. The former are intended to show a maximum fit to the existing data for gapfilling (Sects. 2.2, 3.2). The parameters were optimised simultaneously for $Q_{H,EBC}$ and $Q_{E,EBC}$ with the coefficient of efficiency serving as the objective function (see Eq. (10)) using a SCE-UA (Shuffled Complex Evolution University of Arizona) algorithm (Duan et al., 1992, 1994) followed by a MOSCEM-UA (Multiobjective Shuffled Complex Evolution Metropolis – University of Arizona) algorithm (Vrugt et al., 2003). This procedure was applied to SEWAB just as described in detail by Johnsen et al. (2005). The performance of these runs and their influence on the uncertainty of the upscaled fluxes is discussed in Sect. 4.4.

The realistic parameters were estimated using the detailed information gathered by Kracher et al. (2009). The goal of these model runs is to transfer knowledge of the target land-use, where measurements exist, to an unknown adjacent land-use type by adjusting the site specific parameters (see also Sect. 3.2). As a consequence, parameter optimisation is not appropriate: it is very unlikely to allow estimation of parameters suitable for transferring to other land-use types, even if the physical basis of the model might be correct, due to the problem of equifinality. For this purpose it would be crucial to find a best strategy for adapting model structure and parameters, but this task goes beyond the scope of this work. Only to show in-principle feasibility of this goal, the parameters were taken nearly unchanged from the realistic sets in Kracher et al. (2009) as an "unbiased" estimate, because they did not compare the results with energy balance corrected measurements. An overview of the most important parameter gives Table 1. As A4, like A6, contained maize and showed the same growth pattern during the LITFASS-2003 campaign, both parameter sets differ only slightly in canopy height and roughness length. Nevertheless, some changes were inevitably made to adapt the parameters to their purpose: All hydrological extra modules in SEWAB like ponding, variable infiltration capacity (VIC), subsurface runoff and baseflow

according to the ARNO concept, or depth dependent saturated hydraulic conductivity (Mengelkamp et al., 1999, 2001) are switched off as they all contain parameters for which it is problematic to find realistic values. Furthermore, the roughness lengths are originally assessed from the measurements. In order to exclude this information, z_0 is assumed to be one-tenth of the canopy height. Unfortunately the minimum stomatal resistance $R_{s,min}$, estimated as $300\,\mathrm{s\,m^{-1}}$ for wheat (source: Altman and Dittmer, 1966) and applied for A5, lead to far too low Q_E compared to the observed $Q_{E,EBC}$. While this estimate may be reasonable for an individual *Triticum aestivum* plant, more recent, modelling oriented, studies tend to lower values, e.g. Schulze et al. (1994) give $90\,\mathrm{s\,m^{-1}}$ for cereals (sample size: 5) or Alfieri et al. (2008) with $23\text{--}25\,\mathrm{s\,m^{-1}}$ for two winter wheat fields with intermediate precipitation. Thus, estimates for $R_{s,min}$ differ quite a lot and moreover, Ingwersen et al. (2011) argue for a variable $R_{s,min}$. Therefore we decided to select a uniform $R_{s,min}$ for all sites and choose $60\,\mathrm{s\,m^{-1}}$ for A5, which is the same value as used for maize. And lastly, the LAI value from Kracher et al. (2009) for the maize field (A4 and A6), 1.5, seemed to be estimated for the total surface. Therefore the value was scaled with the inverse fraction of vegetated area 0.7^{-1} for a consistent usage in SEWAB. At the end, the parameters for the sites A4 and A6 differ from A5 only in canopy height, roughness length, albedo, LAI, root depth and fraction of vegetated area.

While these parameters are kept constant in the model for the whole measuring period, the real parameters evolved in time as the experiment was conducted during the growing phase and especially the maize sites showed a significant development of canopy height and, consequently, also other parameters in Table 1. This results in a loss of predictive capacity of the model. On the other hand, a broader range of model performance within the dataset is achieved, raising the opportunity of investigating the relationship between model fit and success of the upscaling procedure.

2.4.3 SEWAB forcing data

The model is forced with a half-hourly dataset derived by the LITFASS-2003 experiment (see Sect. 2.1) consisting of precipitation, air temperature, wind velocity, air pressure, relative humidity, downwelling short-wave and long-wave radiation. As these standard meteorological variables exist for each of the three sites, the optimised runs were driven by the individual forcing sets in order to obtain the best fit. In contrast, for the evaluation of the upscaling scheme the realistic runs were forced with only one dataset (from A5) as would be the case in prospective application. Nevertheless, the different sets were useful for error estimation as discussed in Sect. 4.4.

2.5 Statistics

For evaluation of model performance simple comparisons are carried out using the bias B and the mean absolute error MAE

$$B = N^{-1} \sum_{i=1}^{N} (P_i - O_i) \tag{6}$$

$$\mathrm{MAE} = N^{-1} \sum_{i=1}^{N} |P_i - O_i| \tag{7}$$

with O as the observations and P the model predictions. Thus the MAE provides a measure for model performance or model fitting. There is also a need in this study to quantify the differences between two time series of observations and two series of predictions in the same manner as it is done by the MAE, although the last applications measure dissimilarity more than "errors" in the sense of "failure". In order to avoid confusion with the mean absolute deviation or absolute mean difference, which are commom terms with different meanings, we define the desired measures as

$$D_{\mathrm{obs}} := N^{-1} \sum_{i=1}^{N} \left| O_{1,i} - O_{2,i} \right| \tag{8}$$

$$D_{\mathrm{mod}} := N^{-1} \sum_{i=1}^{N} \left| P_{1,i} - P_{2,i} \right| \tag{9}$$

with O_1, O_2 as observations and P_1, P_2 as predictions from the respective land-use types 1 and 2. As a goodness of fit measure serves the coefficient of efficiency (CoE) or Nash-Sutcliffe coefficient. Introduced by Nash and Sutcliffe (1970), the CoE reads as follows

$$NS = 1 - \frac{\sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2} \tag{10}$$

with $\overline{O}$ as the mean of the observations. The CoE has a range of $[-\infty, 1]$ and should not be mistaken for the more common coefficient of determination R^2. Although the general definitions equal each other, R^2 is predominantly used to assess model performance between observations and predictors via linear regression by ordinary least squares (Everitt, 2002), and in this case partitioning of the total variance into explained and residual variance does hold, restricting the range of R^2 to $[0, 1]$.

2.6 Spatial integration of fluxes

The upscaling scheme proposed in this study and described in Sect. 3.2 connects existing QA/QC tools for EC flux measurements with a procedure for spatial integration of fluxes on a grid size of about 1 km edge length, utilizing footprint and SVAT modelling.

For this task, the relationships between observations, source area and grid representative fluxes should be considered from two perspectives as displayed in Fig. 1. A hypothetical grid cell is assumed to contain M land-use types, each covering a fractional area $a_j^{(g)}$ and featuring the specific fluxes $F_j, j \in [1,2,...,M]$. Typically, measurements are conducted on one land-use of interest or target land-use. We define here the land-use type 1 to be the target land-use and the respective fractional area $a_1^{(g)} = a_{\text{tar}}^{(g)}$ and the target land-use flux $F_1 = F_{\text{tar}}$.

The first perspective to consider is the flux for an entire grid cell (Fig. 1a): Letting the land-use specific fluxes F_j and their fractional areas be known, a grid representative flux $F^{(g)}$ can be obtained by the tile approach,

$$F^{(g)} = \sum_{j=1}^{M} a_j^{(g)} \cdot F_j \tag{11}$$

assuming F_j to be homogeneous for each land-use type (see Fig. 1a). This approach has been already applied to the LITFASS-2003 area on a landscape scale (Beyrich et al., 2006).

On the other hand, observations collected on a tower of a given height reflect the properties of a source region in a corresponding upwind distance from the device. This region can be determined by footprint analysis, delivering a source weight function, which essentially assigns a relative contribution to the measured signal for each pixel in the land-use domain according to the resolution of the footprint model. These relative contributions can be aggregated to a fractional weight for each land-use type $a_j^{(f)}$ (Fig. 1b). Then an observed flux F_{obs} at the tower is related to the land-use specific fluxes, in analogy to Eq. 11, by

$$F_{\text{obs}} = \sum_{j=1}^{M} a_j^{(f)} \cdot F_j \tag{12}$$

The location and extent of this source area is changing rapidly in time, with its temporal dynamics mainly driven by wind direction and stratification in the surface layer.

3 Upscaling of flux measurements

3.1 Representativeness of flux measurements

In this section we aim to provide a framework for assessing a representation error for flux measurements in a heterogeneous landscape. As stated by Nappo et al. (1982), representativeness is a value judgement and depends on situation and purpose. Consequently, we follow Nappo et al. (1982) in general by defining measures for representativeness error in analogy to the proposed criterion for point-to-area representativeness, but the equations are adapted to the specific problem as follows: We restrict ourselves to a situation we judge to be typical when using EC flux measurements for mesoscale modeling or as ground truth for remote sensing data. Let us consider again the hypothetical grid cell, Fig. 1, as described in Sect. 2.6. Assuming each land-use type to be homogeneous, the real flux from the entire grid can be obtained by Eq. (11). Unfortunately, not all land-use specific fluxes F_j required are known in reality, but the measurements provide the flux from areas of target land-use. Although this assumption is often temporally violated, in many cases F_{tar} can be obtained with a certain accuracy by cautious quality checks and footprint analysis followed by a gapfilling algorithm. The most simple, but often used, approach for spatial integration is to assume the target land-use to be representative for the whole grid cell:

$$F_{\text{tar}}^{(g)} = F_{\text{tar}} \tag{13}$$

In contrast, we propose in this study a model-aided grid representative flux $F_{\text{mod}}^{(g)}$:

$$F_{\text{mod}}^{(g)} = a_{\text{tar}}^{(g)} \cdot F_{\text{tar}} + \sum_{j=2}^{M} a_j^{(g)} \cdot F_{j,\text{mod}} \tag{14}$$

This flux is composed of F_{tar} from the measurements and the modelled fluxes $F_{j,\text{mod}}, j \in [2,3,...,M]$ as a surrogate for the missing F_j. Furthermore, substitution of $F^{(g)}$ by the surrogates $F_{\text{mod}}^{(g)}$ or $F_{\text{tar}}^{(g)}$ lead to grid representativeness errors according to Schmid (1997), which can be defined as mean absolute errors δ_{mod} and δ_{tar} respectively.

$$\delta_{\text{mod}} = N^{-1} \sum_{i=1}^{N} \left| F_i^{(g)} - F_{\text{mod},i}^{(g)} \right| \tag{15}$$

$$\delta_{\text{tar}} = N^{-1} \sum_{i=1}^{N} \left| F_i^{(g)} - F_{\text{tar},i}^{(g)} \right| \tag{16}$$

When evaluating the success of the proposed spatial integration $F_{\text{mod}}^{(g)}$, one may ask for the advantage of applying $F_{\text{mod}}^{(g)}$ instead of $F_{\text{tar}}^{(g)}$, i.e. modelling the fluxes from adjacent land-use types and combining it with the target flux instead of simply using F_{tar} for the entire grid cell. This can be formulated as an absolute reduction in error (ARE) or in the style of a proportional reduction in error (PRE):

$$\text{ARE} := \delta_{\text{tar}} - \delta_{\text{mod}} \tag{17}$$

$$\text{PRE} := \frac{\delta_{\text{tar}} - \delta_{\text{mod}}}{\delta_{\text{tar}}} \tag{18}$$

Positive ARE and PRE indicate successful application of the proposed spatial integration (without information regarding significance), negative values indicate that $F_{\text{mod}}^{(g)}$ performs even worse than $F_{\text{tar}}^{(g)}$. Unfortunately ARE and PRE can only be calculated, when $F^{(g)}$ is known, which restricts this evaluation to experimental case studies. Here a case study is constructed with the LITFASS-2003 data set, which is explained in detail in Sect. 4.1.

3.2 Upscaling concept

The general idea of the concept is to derive grid representative fluxes by combining the observations from the target area with modelled data from the other land-use types within the grid cell. The workflow of this concept can be divided into three parts: a flux data post processing part, a gapfilling part and a spatial integration part. Furthermore, the second and the third part depend on the site characteristics. We differentiate here between two situations: The measurement tower is located within the land-use of interest, or target land-use, in a way that a substantial part of the measurements can be attributed as a pure signal from this land-use type. This will be, furthermore, called the *target case*. The opposite is then the *mixed case*, where a pronounced heterogeneity exists around the sensor and most of the data reflects the properties of two or more land-use types. While the target case should be typical for low vegetation landscapes, the mixed case may occur more often for high tower measurements or airborne measurements. With respect to the datasets used, we focus in this study on the first, or target, case.

3.2.1 Target case: identifiable target land-use

In this section the workflow for the target case is described in detail and displayed in Fig. 2. In the first part state-of-the-art flux corrections and quality assessment as used in Mauder et al. (2006) are conducted. A first footprint analysis delivers the spatial context of flux quality and the fraction of flux contribution from the target area in order to characterize the site (Göckede et al., 2008). Consequently, bad rated data is excluded at this step.

For the second part, the preprocessed flux data has to be filtered according to the land-use contributions to ensure that the flux information stems from one land-use type. A high quality standard for this task may be 80 % or higher contribution from the target area. Now the energy balance closure for the target area can be determined and an energy balance closure correction can be applied, which is implemented after Twine

et al. (2000) in this study. After all these steps many data points are discarded, raising the need for gapfilling. Therefore a land surface model is calibrated and validated for the target land-use. Merging this calibrated run with the data yields a full energy balance closed target area flux F_{tar}.

For spatial integration it has to be checked whether fluxes from other land-use types $F_j, j \in [2,3,...,M]$ within the grid cell of interest are expected to deviate considerably from F_{tar}. In this case, the land surface model is activated again to simulate the target land-use with parameter as realistic as possible, yielding $F_{tar,mod}$. This model simulation is then transferred to the other land-use types $F_{j,mod}$ by appropriate modification of the site specific parameters. Now spatial integration can be performed with Eq. (14), delivering $F_{mod}^{(g)}$ as an estimate for the true grid representative flux $F^{(g)}$. As true fluxes from adjacent land-use types are unknown, the PRE, Eq. (18), cannot be evaluated. Instead, a surrogate measure can be obtained quantifying the difference between the modelled fluxes D_{mod}, Eq. (9), and the spatial integration is accepted or rejected according to the threshold X.

$$D_{j,mod} = N^{-1} \sum_{i=1}^{N} \left| F_{j,mod,i} - F_{tar,mod,i} \right| > X \tag{19}$$

The applicability of $D_{j,mod}$ and the threshold value X is discussed in Sect. 4.3. In any case, quality assessment is conducted with respect to point-to-area and target representativeness (Schmid, 1997; Schmid and Lloyd, 1999). A flagging scheme is proposed in Sect. 4.6

3.2.2 Mixed case: pronounced heterogeneity

In this case footprint analysis reveals significant influence of adjacent land-use types on the measured flux and $F_{tar} \neq F_{obs}$. Again considering F_j to be homogeneous for each land-use type and the source weight function from the footprint analysis to be real, F_{obs} is related to the land-use specific fluxes F_j via Eq. (12). The energy balance closure,

5183

however, is difficult to assess for a mixed signal and therefore the simulations of turbulent fluxes tend to overshoot F_{obs}, when applying Eq. (12) to the land-use specific model runs. Recent studies suggest a connection of this imbalance to surface heterogeneity and secondary circulations (Foken, 2008; Foken et al., 2010). Therefore a reasonable assumption is that measurements from nearby surface types exhibit a uniform relative residual of the measurements (i.e. turbulent energy divided by available energy). Under this assumption, the source weight integrated model simulations are proportional to F_{obs}.

$$F_{obs} \sim \sum_{j=1}^{M} a_j^{(f)} \cdot F_{j,mod} \tag{20}$$

Therefore the right hand side of Eq. (20) can be regressed versus F_{obs}. The resulting modelled fluxes are then the basis for the spatial integration.

4 Results and discussion

4.1 Evaluation concept

In Sect. 3.1 a spatial integration strategy is proposed in Eq. (14) with $F_{mod}^{(g)}$, composed of the measured flux from the target land-use F_{tar} and modelled fluxes from other land-use types present in the grid cell intended for upscaling. The advantage (or disadvantage) of this approach over simply using F_{tar} for the entire grid cell is given in terms of absolute and proportional reduction in error ARE and PRE, respectively, defined in Eqs. (17)–(18). These measures can only be given if the true fluxes from all land-use types in the grid cell are known. Therefore, a grid cell is considered here with the postprocessed, energy balance closed and gapfilled flux data from LITFASS-2003 sites A4 (maize), A5 (rye) and A6 (maize), see Sects. 2.1, 2.2, as the true fluxes F_j from the respective surfaces. The realistic SEWAB runs (Sect. 2.4.2) were taken to be the modelled fluxes $F_{j,mod}$, as they could in principle be derived without flux data.

5184

In the case of more than two land-use types within the grid cell ($M > 2$), the PRE and ARE do not only depend on the different fluxes, but also on the respective fractional land-use contributions $a_j^{(g)}$ (use Eqs. (11), (13)–(16) in Eqs. (17)–(18)). Thus, effects originating from flux differences within the cell may be boosted or weakened according to their contribution. In order to focus on the influence of flux differences, we restrict ourselves to only two land-use types to reveal the relationships more clearly. The annotations for $M = 2$ simplify as follows:

$F_{tar} = F_1$ Flux from target land-use

$F_{sur} = F_2$ Flux from surrounding land-use

$a_{tar}^{(g)} = a_1^{(g)}$

$a_{sur}^{(g)} = a_2^{(g)} = 1 - a_{tar}^{(g)}$

$ARE_2 = ARE|_{M=2}$

$PRE_2 = PRE|_{M=2}$

Applying now Eqs. (11), (13)–(16) in Eqs. (17) and (18) for $M = 2$ yields

$$ARE_2 = a_{sur}^{(g)} \cdot N^{-1} \sum_{i=1}^{N} |F_{sur,i} - F_{tar,i}| - |F_{sur,i} - F_{sur,mod,i}| = a_{sur}^{(g)} \cdot (D_{obs} - MAE_{sur}) \tag{21}$$

$$PRE_2 = 1 - \frac{\sum_{i=1}^{N} |F_{sur,mod,i} - F_{sur,i}|}{\sum_{i=1}^{N} |F_{sur,i} - F_{tar,i}|} = 1 - \frac{MAE_{sur}}{D_{obs}} \tag{22}$$

In contrast to the general PRE, the PRE_2 is now independent from fractional land-use areas and the ARE_2 scales only with $1 - a_{tar}^{(g)}$. Simply speaking, the PRE_2 shows positive values and therefore indicates a success in reducing the representativeness

error as long as the difference in observations D_{obs} exceeds the modelling error of the surrounding land-use MAE_{sur}, and vice versa. It should be mentioned here, that the PRE_2 can be interpreted as a generic and baseline-adjusted coefficient of efficiency, as proposed by Legates and McCabe (1999). It differs from the original NS in Eq. (10) by using the sums of absolute errors rather than the sums of squared errors and by replacing $\overline{O}$ with F_{tar} as the so-called baseline series. However, no theoretical distribution exists for PRE_2, which makes it difficult to assess significance, but this problem could be overcome by bootstrapping techniques (Legates and McCabe, 1999).

In this study no significances were assessed. Instead, intercomparisons between the gapfilled and energy balance corrected measurements of turbulent energy fluxes $Q_{H,EBC}$ and $Q_{E,EBC}$ from the LITFASS-2003 stations A4 (maize), A5 (rye) and A6 (maize) and respective SVAT model runs were conducted. In the context of the up-scaling scheme proposed in Sect. 3, these fluxes are assumed to be the real fluxes in a grid cell, containing the target land-use (F_{tar}) and a surrounding land-use (F_{sur}). The threshold values X proposed in Sect. 3.2.1 were established by evaluating daily PRE_2 values in Sect. 4.2 and exploring the relationships to D_{obs} and D_{mod} in 4.3. Finally, the thresholds were discussed by considering the relevant uncertainties in Sect. 4.4 and the influence of the footprint in Sect. 4.5.

4.2 Model performance

Figure 3 displays the energy balance corrected fluxes vs. the calibrated model runs. The Nash-Sutcliffe coefficient NS (Eq. (10)), shows quite reasonable fits for A4 and A5 with values ranging from 0.74 to 0.92. The performance for A4 is worse, which may be attributed to the special conditions of the site influenced by a forest edge at a distance of 150 m. As "special conditions" are common features in field surveys, this data set improves the study by broadening the range of possible outcomes. Compared to the findings from Kracher et al. (2009), this result indicates that the coherence between observations and model outputs can be significally enhanced by energy balance closure correction, and bias can be reduced. However, the application of the spatial

integration requires physically justified parameterisation, and therefore the the fluxes are compared to the realistic model runs as well (Fig. 4). As expected, the realistic runs perform not as well as the calibrated ones, with more bias for all fluxes and poor coherence for A4.

To ensure a consistent averaging for further investigation of daily and summary statistics, all observations are gapfilled with the optimised model runs. While this practice is a necessity for providing average values, some bias might be introduced in the evaluation and evident when comparing model runs with observations, including gapfilled values derived by another model run. Although large gaps exist due to all postprocessing steps (see Sect. 2.2), most of the gaps occur during nighttime. In relation to the set of turbulent flux data, where the absolute values exceed $10\,\mathrm{W\,m^{-2}}$, energy balance corrected flux observations constitute up to 78 % for A4, 67 % for A5 and 53 % for A6. The influence of this procedure is discussed in Sect. 4.4.

Summary statistics according to the evaluation concept (Sect. 4.1) are given in Table 2. Shown are mean fluxes for the whole period as well as mean absolute differences between observations D_{obs}, between simulations of the realistic model runs D_{mod} and the mean absolute error MAE.

The reduction in error measures ARE_2 and the PRE_2 are calculated for each paired combination of the sites A4 to A6, and the indices (1) or (2) indicate whether the first or the second station of a specific pair is the target land-use. The suboptimal performance of the realistic model run A4, as seen already in the scatterplot (Fig. 4), is reproduced in the MAE with highest values for $Q_{\mathrm{H,EBC}}$ and $Q_{\mathrm{E,EBC}}$ at A4 station. Nevertheless, the PRE_2 for A4 as adjacent land-use and A5 as target land-use shows positive values with 0.43 for $Q_{\mathrm{H,EBC}}$ and 0.28 for $Q_{\mathrm{E,EBC}}$, implying a reduction of representativeness error by 43 % and 28 %, respectively. The reason for this is, that the fair model fit of A4 is compensated by an even larger difference between the observations of A4 and A5. So the D_{obs} can be interpreted as the potential: the more the fluxes between land-use types differ, the higher is the representation error of using only the target land-use, and therefore the larger the reduction of representativeness error can be by modelling the

adjacent land-use types. Best PRE_2, ranging from 38 % to 56 %, are achieved with A5 and A6 in the grid cell as well as A4 and A5, provided that A4 is the target land-use. For A4 and A6 in a grid box no significant error reduction can be achieved for all fluxes, and in half of the cases it gets even worse. This outcome is expected, as both A4 and A6 contain maize, showed similar dynamics and therefore differ not so much from each other. Comparing the different fluxes, $Q_{\mathrm{H,EBC}}$ performs best due to better model fitting than $Q_{\mathrm{E,EBC}}$. The absolute reduction in error ARE_2 with $a_{\mathrm{tar}}^{(g)} = 50\,\%$ seem to be quite low, but one has to keep in mind that in this case study the fluxes do not differ so strongly among the sites, anyhow.

As a surrogate for significance, the PRE_2 is also evaluated on a daily basis, and the variation of these daily values are displayed in Fig 5. The specific paired combination and the assignment of the target land-use is coded according to a colour scheme. Most of the results for the whole period (Table 2) are confirmed, with distinct positive PRE_2 for the turbulent fluxes at combinations (A4, A5) and (A5, A6) and not decisive or even negative values for (A4, A6).

4.3 Evaluation of the threshold for spatial integration

The application of the spatial integration includes the modelling of adjacent land-use types, where little or no data is available. The D_{mod} is suggested in Sect. 3.2.1, Eq. (19) as a surrogate for the PRE_2 in order to accept or reject the spatial integration. Its applicability and a reasonable threshold X is discussed in this section, beginning with the D_{obs} and its relationship to the PRE_2.

The D_{obs} exhibit a huge temporal variation on a daily basis (not shown), which is attributed to the temporal evolution of turbulent fluxes due to the rapid growth of the maize fields (A4 and A6). These variations are reflected in daily PRE_2 (see Fig. 5), providing the opportunity to investigate the relationship to D_{obs}, which is displayed in Fig. 6: The scatterplots show daily D_{obs} versus daily PRE_2, the combination of sites again coded according to a colour scheme and the target land-use specified by the

symbols. It is reasonable to assume a coherence between these two variables, because larger differences in observed fluxes raise the potential of successfully replacing an observed flux by a modelled one, a fact which is reflected in PRE_2, where D_{obs} occurs in the denominator (Eq. (22)). In the ideal case, that the daily modelling error for the surrounding land-use MAE_{sur} (numerator in Eq. (22)) is constant, this coherence is a perfect hyperbola $y = 1 - C/x$ with $y \overset{x \to \infty}{\longrightarrow} 1$ and the modelling error equals the D_{obs} between observations for $PRE_2 = 0$. The displayed ideal lines were drawn using the respective MAE_{sur} (see Table 2) for the whole period as constant C. Now deviations from the ideal line are explained only by variations in daily MAE_{sur}. It follows that the outcome of the spatial integration for a whole measurement campaign can be approximated with D_{obs}, and the threshold X is marked by the intersection point of the ideal line with the x-axis (Fig. 6). Using the maximum intersection points for all combinations from Fig. 6, the thresholds for minimal D_{obs} to obtain a positive PRE_2 are 24 W m^{-2} for Q_H and 42 W m^{-2} for Q_E.

The PRE_2, as used until now, gives an answer to the question of whether modelling errors exceed D_{obs}. For applications of the scheme for the target case, however, only F_{tar} is known, and not the fluxes from adjacent areas. Therefore, a surrogate measure has to be found of whether modelling of the adjacent area was successful. A simple attempt would be to relate the difference in mean modelled fluxes D_{mod} to PRE_2 in order to see if a similar threshold of minimum differences can be found as for mean observed fluxes. Figure 7 displays the daily D_{mod} versus PRE_2 in the same style and ideal lines as Fig. 6. It is obvious that the D_{mod} do not reach as high values as the D_{obs}, which is also reflected in the values for the whole period (Table 2). The reason is that the model is not capable of reproducing the whole variance between the fluxes of different sites by changing only a few parameters, as done here (see Table 1). As a consequence, using the thresholds X as established in the previous paragraph for rejecting or accepting the spatial integration according to Eq. (19) leads to only a few false acceptances, indicating the threshold X as a robust estimate. On the other hand, lots of false rejections are produced, so it might be the case in real application that

5189

the spatial integration would be rejected, even though the outcome was positive. One has also to keep in mind that the unknown MAE_{sur} of the adjacent land-use has to be approximated by the model fit of the target land-use MAE_{tar}. In this case study, these estimates are to some extent unbiased as the parameters were estimated for each site simultaneously without fitting, but they differ quite largely from site to site. But it is reasonable to assume that the model fit will get worse when transferring the parameters to an unknown land-use type. So there should be an independent minimum threshold established, which can be derived by uncertainty estimations.

4.4 Consideration of uncertainties

The application of the proposed upscaling scheme includes many steps, which all may introduce errors in the results. The most significant sources of uncertainty are the uncertainty of the flux data itself and uncertainties due to the energy balance closure correction, the gapfilling and the modelling uncertainty of the spatial integration. As the last two sources depend on modelling, these uncertainties can be discussed in terms of model input uncertainty, parameter uncertainty and uncertainty of the model structure. The uncertainty of the flux data is already widely discussed in the community (e.g. Mauder et al., 2007b). Based on the EBEX-2000 EC sensor comparison experiment, Mauder et al. (2006) give uncertainty estimates from 10 to 30 Wm^{-2} for Q_H and 20 to 40 Wm^{-2} for Q_E, depending on sonic anemometer types following the recommendations by Foken and Oncley (1995), and on data quality. Hollinger and Richardson (2005) infer the random error of flux measurements by deviations between two nearby instruments and Richardson et al. (2006) from one time series with subsets under similar conditions. Despite these being substantially different approaches, both yield error quantities similar to Mauder et al. (2006).

4.4.1 Uncertainty due to the energy balance closure and gapfilling

The residual of the measured energy balance, given with about 70 % of the available energy for A5 and A6 and only 60 % for A4, constitutes a large potential source of error. In absolute values, the mean residual (without gapfilling) ranges from 75 to 90 Wm^{-2} during daytime and from −25 to −10 Wm^{-2} during nighttime. Day- and nighttime fluxes were distinguished here by a threshold of 10 Wm^{-2} for R_{swd}, and the data gaps are assumed to be independent from flux magnitude for each group in order to give mean values for non-gapfilled time series. The daytime error is expected to be reduced by the correction for energy balance closure according to the Bowen ratio after Twine et al. (2000), which is seen to be the best first guess for this issue (Foken, 2008), although there is no reason to assume scalar similarity between temperature and water vapour transport (Ruppert et al., 2006; Mauder et al., 2007a), and others propose a correction only for Q_H (Ingwersen et al., 2011). After EBC correction, all missing values are gapfilled, including those where the EBC correction was not applied due to non-compliance with the requirements (see Sect. 2.2). This leads to a mixed influence of gapfilling and EBC correction for daytime values and no influence of EBC correction on the nighttime values, as nearly no data was corrected due to EBC during night.

In order to disentangle the different error sources, one can set up the following assumptions: (i) Although common SVAT models are always a simplification of reality and prone to structural uncertainty, they are in principle able to resemble the true turbulent flux partitioning with a certain accuracy. (ii) The optimisation algorithm used is capable of finding a model solution which approaches this accuracy given the true fluxes. It follows that the gapfilling model would approximate the true fluxes with same accuracy as it approximates the (probably wrong) given $Q_{H,EBC}$ and $Q_{E,EBC}$. Therefore, the daytime gapfilling error can be expressed as MAE and B between the non-gapfilled fluxes and the gapfilling model runs. The turbulent energy fluxes for all stations show maximum MAEs of 36 Wm^{-2} for Q_H and 51 Wm^{-2} for Q_E; for details see Table 3.

For the nighttime case no EBC corrected measurements exist, because the fluxes are either too small or the Bowen ratio is negative. As a consequence, the gapfilling model cannot be calibrated to nighttime conditions and is therefore de facto extrapolated from daytime conditions. Nevertheless, a range can be estimated by comparing model runs, non EBC corrected observations and the residual (see Table 4): the MAE between observations and model runs show only small values within the instrument accuracy. As a consistency check, the sum of the bias B for the turbulent fluxes should approximately equal the residual Res, which is a reasonable expectation when comparing model runs and observations without EBC correction. This is the case at A4 and A6, but almost no bias is visible on A5, although Res equals −26 Wm^{-2}. Therefore, the possible nighttime error has to be in a range of approximately 25 Wm^{-2}.

Daytime data accounts for 66 % and nighttime data for 3 % of the dataset, so the overall gapfilling error can be given as a weighted average: 32 Wm^{-2} and 42 Wm^{-2} for Q_H and Q_E, respectively. In this case study an additional problem arose, that the observation had to be gapfilled with the optimised model runs before comparing with the realistic modelled fluxes, introducing bias to the evaluation. This shortcoming, however, does not severely compromise the results of the evaluation, because the realistic and the optimised model runs' deviation from each other is similar to their difference from the observations: while the mean MAE for all stations and both turbulent fluxes is 39 Wm^{-2} for the unfilled observations and the realistic model and 37 Wm^{-2} for the optimised model, both model runs show a mean D_{mod} for the same data subset of 30 Wm^{-2}. Therefore it can be assumed, that the artificial similarity between modelled data and observations by the use of gapfilling is negligible.

4.4.2 Model uncertainties

The uncertainty of model input is typically related to the measurement errors of the input variables, here precipitation, downward radiation flux components, air temperature, humidity and wind speed. In the specific task of transferring the model to adjacent land-use types, one has to keep in mind that also the input variables may exhibit additional

bias from one land-use to another. The impact on model outcomes was assessed by a Monte Carlo permutation procedure: 10 000 runs were conducted for each station with the realistic parameters remaining fixed and the model input being varied: for each run a random deviation from the original input was drawn – independently for each variable – from a normal distribution, and this deviation was added as constant bias for the respective run. The standard deviations of the normal distribution are estimated with the help of the existing forcing data sets from all three stations and set to 0.5 K for air temperature, 3 % for relative humidity, $0.3\,\mathrm{m\,s}^{-1}$ for u and $10\,\mathrm{W\,m}^{-2}$ for R_{swd} and R_{lwd}. As the modulation of precipitation makes only sense for the time steps when precipitation occurs, a constant multiplier is used instead of a constant offset, again from a normal distribution with a mean of 1 and a standard deviation of 0.2. The distribution of the differences between the resulting model runs and the reference run (unaltered forcing data) show a distinct diurnal cycle and a temporal variation, with larger deviations after rain events. The mean diurnal deviations are displayed as median and specific quantiles in Fig. 8 for A5. Beside the diurnal cycle one can see that the model behaviour is especially sensitive in the morning and evening hours in case of $Q_{E,EBC}$. Nevertheless, on a 95 % confidence level, the mean differences to the reference run range from -16 to $20\,\mathrm{W\,m}^{-2}$ for Q_H, from -10 to $6\,\mathrm{W\,m}^{-2}$ for Q_E. This uncertainty may modulate the model results among land-use types in both directions, either enhancing or smoothing the flux differences between land-use types.

The assessment of model parameter uncertainty has been attracting great interest within the scientific community and many methods already exist, accounting for parameter uncertainty and other model-related uncertainties. A comparison between the most popular methods can be found in e.g. Yang et al. (2008). However, these methods fail for the application in this study, because they rely on parameter optimisation, which is not the case when the parameters are fixed to physically reasonable values, which are then transferred to the site specific conditions of other land-use types. On the other hand, the uncertainty arising from this parameter estimation strategy is directly reflected in the evaluation of model performance with respect to the PRE_2, as is done

in Sect. 4.3. It is therefore argued that the parameter uncertainty has to be within the maximum MAE given in Table 2: $24\,\mathrm{W\,m}^{-2}$ for Q_H and $42\,\mathrm{W\,m}^{-2}$ for Q_E. As this estimation is not fully conclusive, more effort has to be put into the evaluation of the parameter estimation strategy in future studies. A necessary prerequisite for future application of the scheme will be a sensitivity analysis of the used model, in order to assess whether the differences between the model runs are able to exceed the uncertainty for a given range of the realistic parameter.

A critical issue is the uncertainty in model structure. An incorrect model structure may even lead to opposite tendencies when adjusting the site specific parameters to the unknown land-use type, which cannot be controlled without measurements. Especially in this case study, a serious drawback originates from the model failure for A5 when using the minimum stomatal resistance as estimated a priori. Here the knowledge to set this parameter equal to the other sites was deduced from the data only. Therefore the spatial integration procedure is only applicable to those land-use types for which prior ecological knowledge from comparable studies is available in order to assess whether the model outcome is reasonable or not. A quantitative assessment of structural uncertainty can only be given via model ensemble approaches, but such an investigation is beyond the scope of this paper.

4.4.3 Summary uncertainty estimation

In Sect. 4.3, a threshold for a minimal D_{mod} was derived, with $24\,\mathrm{W\,m}^{-2}$ for Q_H and $42\,\mathrm{W\,m}^{-2}$ for Q_E in this case study, for the application of the spatial integration. This has to be refined with the findings from the uncertainty estimation. A summary of the uncertainties considered is given in Table 5. These different error sources interact in a complicated way, and the application of a linear error propagation model seems useless. Nevertheless, one can derive the necessary condition, that the difference between modelled fluxes D_{mod} should exceed at least the largest uncertainty. From Table 5 it follows that a meaningful estimate for X would be $35\,\mathrm{W\,m}^{-2}$ for Q_H and $50\,\mathrm{W\,m}^{-2}$ for Q_E.

4.5 Influence of the footprint

The representativeness of measurements within a heterogeneous landscape has been investigated in the previous sections so far without taking the footprint into account. Thus F_{obs} was simply assumed to equal F_{tar}. As already mentioned in Sect. 3.2.2, there are situations where other land-use types are situated so close to the device, that footprint analysis detects significant influence on the measured signal from these areas. In the context of this study, the question arises of whether SVAT modelling might be helpful for disentangling potentially mixed observed signals as suggested with Eq. (20). Size and location of the footprint area change quickly in time, determined by wind direction and atmospheric stability. Although measurement height and roughness length define the footprint area as well, these variables are fixed or do not vary very quickly, thus they have hardly any impact on the temporal variations. On the other hand, flux magnitude is also connected to atmospheric stability. Large turbulent fluxes occur typically during daytime under neutral or unstable conditions, while during nighttime a stable boundary layer develops and the fluxes were generally low. In summary, flux magnitude and footprint size are related to each other, exhibiting high fluxes from small footprints and low fluxes from large footprints. As a simple example, flux observations are considered from A5, measured on a virtual tower at 5 m height, and the surrounding land-use A4 appears at a distance of 50 m from the device. Unstable and neutral conditions occur for ca. 75 % of the used data points, while 25 % were sampled during stable stratification. Footprint analysis in this virtual setup shows that for unstable and neutral conditions the properties of the target land-use are well reflected, with contributions higher than 89 %. Significant influence of the surrounding land-use occurs only under stable stratification, with target land-use contributions ranging from 40 % to 89 %. Along with flux magnitude, also the absolute difference between the observations of A4 and A5 depend on stratification, which is illustrated in Fig. 9. The difference is plotted against the contribution of the target land-use (here A5) and the individual data points were distinguished for different stratification regimes.

Boxplots represent the data in classes of contribution (5 % width), with the number of data points within a class below the boxes. This shows clearly that all fluxes deviate considerably only during unstable and neutral conditions, while under stable stratification the differences turn out to be quite small for Q_H and Q_E. Therefore, it is concluded that under stable stratification, the representativeness of the observations for the target land-use cannot be enhanced by footprint analysis according to Eq. (20), even if significant contributions of other land-use types occur. Considering the error sources of the measurements, footprint analysis and SVAT modelling, F_{tar} is reflected by F_{obs} quite reasonably under such conditions. This approach is only useful when significant influence from different land-use types is detected under neutral and unstable conditions, or if stratification regimes deviate substantially among land-use types. Such a situation can be given with measurements directly at the edge of two land-use types or with higher tower measurements within the surface layer.

4.6 Quality flagging for representativeness

According to the upscaling scheme presented in Sect. 3.2, the processed data will be classified with respect to spatial representativeness for the grid cell. For the target case it can be assumed that most of the measurements reflect the target land-use, thus the sensor location bias, including the individual footprints, is of little importance. Such a flagging scheme will depend on the fraction of the target land-use within the grid $a_{tar}^{(g)}$, the measurement error and on the relation of D_{mod} to the threshold X (Eq. (19)), defining whether additional modelling should be applied for spatial integration. For $a_{tar}^{(g)} > 0.8$, representativeness can be assumed within the given accuracy in any case, but with a decreasing $a_{tar}^{(g)}$, the differences in fluxes between the target land-use and other surfaces get more and more important. In order to derive a feasible quality scheme, the mentioned variables were classified in advance as given in Table 6. The flags for the measurement error (MFlag) use the classification from Table 5, but the last threshold is extended to also consider the error of gapfilling. D_{mod} is classified with only two

limits, the latter equals the threshold established in Sect. 4.4 for applying the spatial integration (HFlag). Thus HFlag = 3 indicates that the measurements were completed with the model runs for the enire grid cell. As this threshold was derived as the error of the spatial integration, HFlag = 2 and 3 are assumed to exhibit the same error for calculation of the total error for the grid representative flux. The respective overall flag ranges from 1 to 5, corresponding to grid flux errors of 20–45 Wm^{-2} and more for Q_H and 30–60 Wm^{-2} and more for Q_E (Table 7). These errors were calculated by using the limits from Table 6, weighted by the target land-use contribution $a_{tar}^{(g)}$ and the surrounding land-use $1 - a_{tar}^{(g)}$. Two problems restrict the calculation: (i) the EBC error is unknown and therefore neglected, and (ii) D_{mod} is used as a surrogate for the unknown D_{obs}, which is systematically higher than D_{mod} as discussed in Sect. 4.3. Therefore, to match at least the conditions of this case study, the error for heterogeneity is expanded by a regression between D_{obs} and D_{mod} from the whole period (Table 2), i.e. D_{mod} is multiplied by 1.2 and a offset of 12 Wm^{-2} was added. Considering the whole measurement period as a basis, the spatial integration would be rejected in any combination of the sites due to small D_{mod} with respect to X.

5 Conclusions

In this study an upscaling scheme to enhance representativeness of EC flux measurements for mesoscale modelling and evaluation of remote sensing data has been proposed. In order to evaluate applicability and the work flow of the scheme, EC flux measurements (Q_H and Q_E) of the LITFASS-2003 experiment from three adjacent stations have been compared with SVAT model runs of the model SEWAB. Model performance was assessed with simple statistical measures and related to differences in observed fluxes. Furthermore, the significant error sources were discussed individually in order to estimate the feasibility of the upscaling scheme and to identify a reasonable threshold for applying the spatial integration. The influence of the footprint was discussed

concerning the representativeness of the observations for the target land-use and finally, a classification for the quality of grid representative fluxes was proposed.

It has been shown that the upscaling scheme is able to enhance representativeness of fluxes on a grid scale in general, as the model runs yield a reduction in the representation error for the turbulent energy fluxes ranging from 28 up to 56 %, when evaluating a maize field together with a rye field in a grid cell. The absolute reduction, however, is quite low due to a general small difference between the fluxes of the two LITFASS-2003 sites for the whole period. The evaluation on a daily basis showed that the sites differ more in temporal dynamics than in mean fluxes and the reduction in representation error was, to a large extent, achieved by an enhanced representation of this dynamics. The success of applying the spatial integration has been related to mean absolute deviations between model runs D_{mod} of adjacent land-use types. Using D_{mod} as a surrogate measure to decide about model success in real application, minimal thresholds have been established with 25 $W\,m^{-2}$ for Q_H and 40 $W\,m^{-2}$ for Q_E. When restricting the model runs for the spatial integration to only few parameters to be transferred, these thresholds become more robust, but on the other hand the ability of the model to capture the variance of the observations is strongly reduced. Other important sources of uncertainty were examined separately. When using the largest error as a minimum benchmark, the thresholds X have to be refined to 35 $W\,m^{-2}$ for Q_H and 50 $W\,m^{-2}$ for Q_E. Under the assumption that similar model errors can be achieved elsewhere, these thresholds can be extrapolated to other study sites. Furthermore, it is shown that the flux magnitude depends on stability to a large extent and is therefore strongly connected with the size of the footprint, leading to large footprints, especially when the fluxes are small. Although this is well known, it helps to assess the impact of the adjacent land-use: most pronounced contributions of the surrounding land-use happen during stable stratification, when the fluxes from different land-use types show only small deviations. The target representativeness of observations is therefore not affected as long as no significant influence comes from adjacent land-use types under neutral and unstable conditions. These considerations should be taken into account

whenever target representativeness is assessed, e.g. with the sensor location bias according to Schmid (1997). The proposed quality scheme for representativeness combines measurement errors of the target area with the heterogeneity of the whole grid cell and considers model uncertainties with respect to gapfilling and modelling
5 the adjacent land-use types. It provides the data user with a maximum error for grid representative turbulent fluxes regardless of the error sources, which can be utilized to infer the uncertainty of mesoscale models and satellite data based on the ground observations. Although designed for the case of a target land-use type within the grid, the scheme could, in principle, be transferred to the mixed case (Sect. 3.2.2).

10 The validation of model performance for adjacent land-use types, where typically no flux measurements are available, remains a major issue. The D_{mod} from the target land-use (where measurements exist) and another land-use has been proposed as a surrogate measure for D_{obs}. But as the realistic model runs show substantially lower differences between land-use types than the observations, the spatial integration for
15 the whole period would be rejected in any case for the given sites, even if the PRE_2 showed a successful outcome. Therefore, more effort has to be put into developing a "best practice" to utilize a SVAT model for this task. Some problems remain unsolved within this study. One is how to tackle the structural uncertainty for the model, which is a crucial point when using the model "blind". From this study it can be concluded that
20 additional knowledge has to be acquired for the unknown land-use. This could be either some point flux measurements or a profound "expert knowledge" for the given surface, that makes a user able to rank the model results with their expectation. On the other hand, one can never be sure to get the right fluxes without any observations. Therefore it is recommended to stick to the target case only, when minor influences from different
25 surfaces are apparent. Otherwise sensor installation on a surface edge, sampling over two surface types or occasionally moving the instruments should be preferred. As well, the uncertainty due to the EBC correction remains unclear – a very critical issue – with the potential error to be made being in the magnitude of the residual. But further steps are therefore needed towards a reasonable closing of the observed energy balance

5199

gap. As the aim of this study was to show the applicability of the proposed upscaling strategy in general, this task is beyond the scope of this paper; even so, achievements in this issue would strengthen the impact of the work presented herein. Deriving a closed energy balance is a necessity for upscaling turbulent fluxes, no matter whether
5 they were used for validation of MODIS data or in mesoscale model applications.

Appendix A

List of abbreviations and symbols.

CoE	Coefficient of efficiency	
EBC	Energy balance closure	
EC	Eddy covariance	
LITFASS	**L**indenberg **I**nhomogeneous **T**errain–**F**luxes between **A**tmosphere and **S**urface: a long term **S**tudy	
SEWAB	SVAT scheme: Surface Energy and Water Balance	
SVAT	Soil – Vegetation – Atmosphere – Transport scheme	
$a_j^{(f)}$	Contribution of the land-use area j to the measured flux as a fraction of the source area, determined by footprint calculation	–
$a_{tar}^{(f)}$	Contribution of the target land-use to the measured flux as a fraction of the source area, determined by footprint calculation	–
$a_j^{(g)}$	Contribution of the land-use area j to the grid representative flux as a fraction of the grid area	–
$a_{tar}^{(g)}$	Contribution of the target land-use to the grid representative flux as a fraction of the grid area	–
ARE	Absolute reduction in error	*
B	Bias between two series of fluxes	*
Bo	Bowen ratio	–
C_D	Drag coefficient	–

5200

C_H	Stanton number	–
c_p	Specific heat capacity at constant pressure	$\mathrm{J\,kg^{-1}\,K^{-1}}$
D_{mod}	Mean absolute difference between two time series of predictions	*
D_{obs}	Mean absolute difference between two time series of observations	*
$F^{(g)}$	Representative flux for the entire grid cell	*
$F^{(g)}_{mod}$	Grid representative flux, composed from the target flux and the modelled flux from the surrounding area	*
$F^{(g)}_{tar}$	Grid representative flux, composed solely from the target flux	*
F_j	Flux from land-use j	*
F_{obs}	Measured flux	*
$F_{sur,mod}$	Modelled flux from the surrounding area	*
F_{sur}	Flux from the surrounding area (in case of a grid cell containing only two types of land-use)	*
F_{tar}	Flux from the target area	*
M	Number of land-use types within a grid cell	–
MAE	Mean absolute error	*
N	Length of a time series	–
NS	Coefficient of efficiency, original Nash-Sutcliffe coefficient	–
O	Observations (flux or scalar)	*
P	Predictions (flux or scalar)	*
PRE	Proportional reduction in error	–
$Q_{E,EBC}$	Observed latent heat flux, energy balance corrected	$\mathrm{W\,m^{-2}}$
Q_E	Latent heat flux, evapotranspiration	$\mathrm{W\,m^{-2}}$
Q_G	Ground heat flux	$\mathrm{W\,m^{-2}}$
$Q_{H,EBC}$	Observed sensible heat flux, energy balance corrected	$\mathrm{W\,m^{-2}}$
Q_H	Sensible heat flux	$\mathrm{W\,m^{-2}}$
Q_R^*	Net radiation	$\mathrm{W\,m^{-2}}$
R_{lwd}	Downwelling long-wave radiation	$\mathrm{W\,m^{-2}}$
R_{swd}	Downwelling short-wave radiation	$\mathrm{W\,m^{-2}}$

Res	Observed residual of the experimental energy budget	$\mathrm{W\,m^{-2}}$
T	Absolute temperature	K
T_g	Absolute temperature at the surface	K
u	Wind velocity (alongwind component)	$\mathrm{m\,s^{-1}}$
u_*	Friction velocity	$\mathrm{m\,s^{-1}}$
z	Height	m
z_m	Measurement height	m
δ_{mod}	Mean absolute error of approximating $F^{(g)}$ with $F^{(g)}_{mod}$	*
δ_{tar}	Mean absolute error of approximating $F^{(g)}$ with F_{tar}	*
ε	Emissivity	–
λ	Heat of evaporation for water	$\mathrm{J\,kg^{-1}}$
σ	Stefan-Boltzmann constant	$\mathrm{W\,m^{-2}\,K^{-4}}$

* The dimension is not defined, as the variable represents different quantities

References

Alfieri, J. G., Niyogi, D., Blanken, P. D., Chen, F., LeMone, M. A., Mitchell, K. E., Ek, M. B., and Kumar, A.: Estimation of the Minimum Canopy Resistance for Croplands and Grasslands Using Data from the 2002 International H2O Project, Mon. Weather. Rev., 136, 4452–4469, doi:10.1175/2008MWR2524.1, 2008. 5176

Altman, P. L. and Dittmer, D. S.: Environmental Biology, Federation of American Societies for Experimental Biology, Bethesda, Maryland, 1966. 5176

Avissar, R.: Scaling of land-atmosphere interactions: An atmospheric modelling perspective, Hydrol. Process., 9, 679–695, doi:10.1002/hyp.3360090514, 1995. 5169

Avissar, R. and Pielke, R. A.: A parameterization of heterogeneous land surfaces for atmospheric numerical-models and its impact on regional meteorology, Mon. Weather. Rev., 117, 2113–2136, 1989. 5168

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., Paw, K. T., Pilegaard, K., Schmid, H. P., 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, B. Am. Meteorol. Soc., 82, 2415–2434, 2001. 5167

Baldocchi, D., Krebs, T., and Leclerc, M.: "Wet/dry Daisyworld": a conceptual tool for quantifying the spatial scaling of heterogeneous landscapes and its impact on the subgrid variability of energy fluxes, Tellus B, 57, 175–188, 2005. 5169

Best, M. J., Beljaars, A., Polcher, J., and Viterbo, P.: A proposed structure for coupling tiled surfaces with the planetary boundary layer, J. Hydrometeorol., 5, 1271–1278, 2004. 5168

Beyrich, F. and Adam, W.: Site and Data Report for the Lindenberg Reference Site in CEOP – Phase1, 230, Berichte des Deutschen Wetterdienstes, 55 pp., 2007. 5170

Beyrich, F. and Mengelkamp, H. T.: Evaporation over a Heterogeneous Land Surface: EVA GRIPS and the LITFASS-2003 Experiment – An Overview, Bound.-Lay. Meteorol., 121, 5–32, 2006. 5170

Beyrich, F., Leps, J.-P., Mauder, M., Bange, J., Foken, T., Huneke, S., Lohse, H., Lüdi, A., Meijninger, W., 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, 2006. 5179

Chen, B., Black, T., Coops, N., Hilker, T., Trofymow, J., and Morgenstern, K.: Assessing tower flux footprint climatology and scaling between remotely sensed and eddy covariance measurements, Bound.-Lay. Meteorol., 130, 137–167, doi:10.1007/s10546-008-9339-1, 2009. 5167, 5168, 5169

Chen, B., Ge, Q., Fu, D., Yu, G., Sun, X., Wang, S., and Wang, H.: A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling, Biogeosciences, 7, 2943–2958, doi:10.5194/bg-7-2943-2010, 2010. 5168

5203

Clapp, R. B. and Hornberger, G. M.: Empirical Equations for some Soil Hydraulic Properties, Water. Resour. Res., 14, 601–604, 1978. 5174, 5210

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water. Resour. Res., 28, 1015–1031, 1992. 5175

Duan, Q., Sorooshian, S., and Gupta, V.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, 1994. 5175

Everitt, B.: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge, UK, 2nd edn., 2002. 5178

Finn, D., Lamb, B., Leclerc, M. Y., and Horst, T. W.: Experimental evaluation of analytical and Lagrangian surface-layer flux footprint models, Bound.-Lay. Meteorol., 80, 283–308, 1996. 5169

Foken, T.: The energy balance closure problem: An overview, Ecol. Appl., 18, 1351–1367, 2008. 5172, 5184, 5191

Foken, T. and Leclerc, M. Y.: Methods and limitations in validation of footprint models, Agr. Forest. Meteorol., 127, 223–234, 2004. 5168, 5169

Foken, T. and Oncley, S.: Workshop on instrumental and methodical problems of land surface flux measurements, B. Am. Meteorol. Soc., 76, 1191–1193, 1995. 5190, 5214

Foken, T., Göckede, M., Mauder, M., Mahrt, L., Amiro, B., and Munger, J.: Post-field data quality control, in: Handbook of micrometeorology: A guide for surface flux measurement and analysis, edited by: Lee, X., Massman, W., and Law, B., 181–208, Kluwer, Dordrecht, 2004. 5171, 5214

Foken, T., Mauder, M., Liebethal, C., Wimmer, F., Beyrich, F., Leps, J.-P., Raasch, S., DeBruin, H., Meijninger, W., and Bange, J.: Energy balance closure for the LITFASS-2003 experiment, Theor. Appl. Climatol., 101, 149–160, doi:10.1007/s00704-009-0216-8, 2010. 5171, 5184

Göckede, M., Rebmann, C., and Foken, T.: A combination of quality assessment tools for eddy covariance measurements with footprint modelling for the characterisation of complex sites, Agr. Forest. Meteorol., 127, 175–188, 2004. 5169, 5173

Göckede, M., Markkanen, T., Mauder, M., Arnold, K., Leps, J. P., and Foken, T.: Validation of footprint models using natural tracer measurements from a field experiment, Agr. Forest. Meteorol., 135, 314–325, 2005. 5168, 5172, 5173

Göckede, M., Markkanen, T., Hasager, C. B., and Foken, T.: Update of a footprint-based approach for the characterisation of complex measurement sites, Bound.-Lay. Meteorol., 118, 635–655, 2006. 5169

5204

Göckede, M., Foken, T., Aubinet, M., Aurela, M., Banza, J., Bernhofer, C., Bonnefond, J. M., Brunet, Y., Carrara, A., Clement, R., Dellwik, E., Elbers, J., Eugster, W., Fuhrer, J., Granier, A., Grunwald, T., Heinesch, B., Janssens, I. A., Knohl, A., Koeble, R., Laurila, T., Longdoz, B., Manca, G., Marek, M., Markkanen, T., Mateus, J., Matteucci, G., Mauder, M., Migli-avacca, M., Minerbi, S., Moncrieff, J., Montagnani, L., Moors, E., Ourcival, J. M., Papale, D., Pereira, J., Pilegaard, K., Pita, G., Rambal, S., Rebmann, C., Rodrigues, A., Rotenberg, E., Sanz, M. J., Sedlak, P., Seufert, G., Siebicke, L., Soussana, J. F., Valentini, R., Vesala, T., Verbeeck, H., and Yakir, D.: Quality control of CarboEurope flux data - Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems, Biogeosciences, 5, 433–450, doi:10.5194/bg-5-433-2008, 2008. 5169, 5171, 5173, 5182

Hasager, C. B. and Jensen, N. O.: Surface-flux aggregation in heterogeneous terrain, Q. J. Roy. Meteor. Soc., 125, 2075–2102, 1999. 5168

Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree. Physiol., 25, 873–885, 2005. 5190

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, 2009. 5169

Ingwersen, J., Steffens, K., Högy, P., Warrach-Sagi, K., Zhunusbayeva, D., Poltoradnev, M., Gäbler, R., Wizemann, H.-D., Fangmeier, A., Wulfmeyer, V., and Streck, T.: Comparison of Noah simulations with eddy covariance and soil water measurements at a winter wheat stand, Agr. Forest. Meteorol., 151, 345–355, doi:10.1016/j.agrformet.2010.11.010, 2011. 5176, 5191

Johnsen, K.-P., Mengelkamp, H.-T., and Huneke, S.: Multi-objective calibration of the land surface scheme TERRA/LM using LITFASS-2003 data, Hydrol. Earth. Syst. Sc., 9, 586–596, doi:10.5194/hess-9-586-2005, 2005. 5175

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a bio-sphere model, Biogeosciences, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009. 5167

Kim, J., Guo, Q., Baldocchi, D., Leclerc, M., Xu, L., and Schmid, H.: Upscaling fluxes from tower to landscape: Overlaying flux footprints on high-resolution (IKONOS) images of veg-etation cover, Agr. Forest. Meteorol., 136, 132–146, doi:10.1016/j.agrformet.2004.11.015, 2006. 5169

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, 2002. 5168

Kracher, D., Mengelkamp, H. T., and Foken, T.: The residual of the energy balance closure and its influence on the results of three SVAT models, Meteorol. Z., 18, 1–15, 2009. 5172, 5173, 5175, 5176, 5186, 5210

Legates, D. R. and McCabe, G. J.: Evaluating the use of "goodness-of-fit" measures in hydro-logic and hydroclimatic model validation, Water. Resour. Res., 35, 233–241, 1999. 5186

Louis, J.: A parametric model of vertical eddy fluxes in the atmosphere, Bound.-Lay. Meteorol., 17, 187–202, 1979. 5174

Ma, Y., Zhong, L., Su, Z., Ishikawa, H., Menenti, M., and Koike, T.: Determination of re-gional distributions and seasonal variations of land surface heat fluxes from Landsat-7 En-hanced Thematic Mapper data over the central Tibetan Plateau area, J. Geophys. Res., 111, D10305, doi:10.1029/2005JD006742, 2006. 5168

Markkanen, T., Steinfeld, G., Kljun, N., Raasch, S., and Foken, T.: Comparison of conventional Lagrangian stochastic footprint models against LES driven footprint estimates, Atmos. Chem. Phys., 9, 5575–5586, doi:10.5194/acp-9-5575-2009, 2009. 5169, 5173

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, 1–9, 2010. 5173

Mason, P.: The formation of areally-averaged roughness length, Q. J. Roy. Meteor. Soc., 114, 399–420, 1988. 5167

Mauder, M. and Foken, T.: Documentation and instruction manual of the eddy covariance soft-ware package TK2, Work Report University of Bayreuth, Dept. of Micrometeorology, ISSN 1614-8916, 2004. 5171

Mauder, M., Liebethal, C., Gockede, M., Leps, J. P., Beyrich, F., and Foken, T.: Processing and quality control of flux data during LITFASS-2003, Bound.-Lay. Meteorol., 121, 67–88, 2006. 5171, 5182, 5190, 5214

Mauder, M., Desjardins, R. L., and MacPherson, I.: Scale analysis of airborne flux measure-ments over heterogeneous terrain in a boreal ecosystem, J. Geophys. Res., 112, D13112, doi:10.1029/2006JD008133, 2007a. 5191

Mauder, M., Oncley, S., Vogt, R., Weidinger, T., Ribeiro, L., Bernhofer, C., Foken, T., Kohsiek, W., De Bruin, H., and Liu, H.: The energy balance experiment EBEX-2000. Part II: Intercom-parison of eddy-covariance sensors and post-field data processing methods, Bound.-Lay. Meteorol., 123, 29–54, doi:10.1007/s10546-006-9139-4, 2007b. 5190

Mengelkamp, H. T., Warrach, K., and Raschke, E.: SEWAB – a parameterization of the Surface Energy and Water Balance for atmospheric and hydrologic models, Adv. Water. Resour., 23, 165–175, 1999. 5173, 5176

Mengelkamp, H. T., Kiely, G., and Warrach, K.: Evaluation of the hydrological components added to an atmospheric land-surface scheme, Theor. Appl. Climatol., 69, 199–212, 2001. 5176

Mengelkamp, H. T., Beyrich, F., Heinemann, G., Ament, F., Bange, J., Berger, F., Bösenberg, J., Foken, T., Hennemuth, B., Heret, C., Huneke, S., Johnsen, K. P., Kerschgens, M., Kohsiek, W., Leps, J. P., Liebethal, C., Lohse, H., Mauder, M., Meijninger, W., Raasch, S., Simmer, C., Spieß, T., Tittebrand, A., Uhlenbrook, S., and Zittel, P.: Evaporation over a heterogeneous land surface: The EVA-GRIPS project, B. Am. Meteorol. Soc., 87, 775–786, 2006. 5170

Metzger, S., Ma, Y., Markkanen, T., Göckede, M., Li, M., and Foken, T.: Quality assessment of Tibetan Plateau eddy covariance measurements utilizing footprint modeling, Advances in Earth Science, 21, 1260–1267, 2006. 5169

Mölders, N., Raabe, A., and Tetzlaff, G.: A comparison of two strategies on land surface heterogeneity used in a mesoscale β meteorological model, Tellus A, 48, 733–749, doi:10.1034/j.1600-0870.1996.00012.x, 1996. 5168

Nappo, C., Caneill, J. Y., Furman, R. W., Gifford, F. A., Kaimal, J. C., Kramer, M. L., Lockhart, T. J., Pendergast, M. M., Pielke, R. A., Randerson, D., Shreffler, J. H., and Wyngaard, J. C.: The Workshop on the Representativeness of Meteorological Observations, June 1981, Boulder, Colo., B. Am. Meteorol. Soc., 63, 761–764, 1982. 5180

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970. 5178

Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, Mon. Weather. Rev., 117, 536–549, 1989. 5174

Rannik, U., Markkanen, T., Raittila, J., Hari, P., and Vesala, T.: Turbulence statistics inside and over forest: Influence on footprint prediction, Bound.-Lay. Meteorol., 109, 163–189, 2003. 5172

Raupach, M. R. and Finnigan, J. J.: Scale issues in boundary-layer meteorology: Surface energy balances in heterogeneous terrain, Hydrol. Process., 9, 589–612, doi:10.1002/hyp.3360090509, 1995. 5169

Rebmann, C., Göckede, M., Foken, T., Aubinet, M., Aurela, M., Berbigier, P., Bernhofer, C., Buchmann, N., Carrara, A., Cescatti, A., Ceulemans, R., Clement, R., Elbers, J. A., Granier,

T., and Leclerc, M. Y.: Flux and concentration footprint modelling: State of the art, Environ. Pollut., 152, 653–666, 2008. 5168

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water. Resour. Res., 39, 1214, doi:10.1029/2002WR001746, 2003. 5175

Yang, J., Reichert, P., Abbaspour, K., Xia, J., and Yang, H.: Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, J. Hydrol., 358, 1–23, doi:10.1016/j.jhydrol.2008.05.012, 2008. 5193

Table 1. Most important parameter for the realistic model runs. Parameters, which differ among sites, are: albedo a, roughness length z_0, canopy height h_c, fraction of vegetated area f_{veg}, rooting depth z_r and leaf area index LAI. The other parameters were equal for all sites: emissivity ε, minimum stomatal resistance $R_{s,min}$, maximum stomatal resistance $R_{s,max}$, radiation limit value for stomatal resistance R_{GL}, saturated hydraulic conductivity K_{sat}, matrix potential at saturation Ψ_{sat}, porosity η_{sat}, volumetric water content at field capacity η_{FC}, volumetric water content at wilting point η_{WP}, exponent b for relationships after Clapp and Hornberger (1978).

Parameter	Unit	A4	A5	A6
a	–	0.20	0.18	0.20
ε	–		0.986	
z_0	m	0.065[*]	0.120[*]	0.045[*]
h_c	m	0.65	1.20	0.45
f_{veg}	–	0.7	1.0	0.7
z_r	m	0.25	0.60	0.25
LAI	–	2.14[*]	2.73	2.14[*]
$R_{s,min}$	sm^{-1}	60	60[*]	60
$R_{s,max}$	sm^{-1}		2500	
R_{GL}	Wm^{-2}		90	
K_{sat}	ms^{-1}		5.23×10^{-6}	
Ψ_{sat}	m		−0.1	
η_{sat}	m^{3}m^{-3}		0.44	
η_{FC}	m^{3}m^{-3}		0.342	
η_{WP}	m^{3}m^{-3}		0.15	
b	–		4.8	

[*] Parameter differs from Kracher et al. (2009), see Sect. 2.4.2.

Table 2. Mean fluxes and selected variables for the whole period, according to the evaluation concept in Sect. 4.1. The absolute reduction of the representativeness error (ARE_2) scales with $1 - a_{tar}^{(g)}$, and the presented values are valid for a fractional target area $a_{tar}^{(g)} = 70\%$. The ARE_2 and the PRE_2 are calculated for each paired combination of the sites A4 to A6, and the indices (tar = 1) or (tar = 2) indicate whether station (1) or station (2) of a specific pair is the target land-use.

	A4 (maize)		A5 (rye)		A6 (maize)	
	Q_H [Wm^{-2}]	Q_E [Wm^{-2}]	Q_H [Wm^{-2}]	Q_E [Wm^{-2}]	Q_H [Wm^{-2}]	Q_E [Wm^{-2}]
Observed mean	66.6	75.5	45.4	113.4	65.0	71.4
Modelled mean	71.5	58.7	57.9	85.0	64.0	65.7
MAE	24.0	41.7	18.6	30.1	18.8	26.3
	A4 (1) ↔ A5 (2)		A5 (1) ↔ A6 (2)		A6 (1) ↔ A4 (2)	
D_{obs}	42.2	57.6	34.6	48.8	19.0	28.0
D_{mod}	26.0	36.3	22.8	30.9	8.2	7.5
$ARE_2(\text{tar}=1)\big|_{a_{tar}^{(g)}=50\%}$	11.6	13.8	7.9	11.2	−2.5	−6.9
$ARE_2(\text{tar}=2)\big|_{a_{tar}^{(g)}=50\%}$	9.1	7.9	8.0	9.4	0.1	0.8
	[−]	[−]	[−]	[−]	[−]	[−]
$PRE_2(\text{tar}=1)$	0.56	0.48	0.46	0.46	−0.26	−0.49
$PRE_2(\text{tar}=2)$	0.43	0.28	0.46	0.38	0.01	0.06

Table 3. MAE and bias B between the optimised model runs and the unfilled – but energy balance closure corrected – observations during daytime and the daytime residual Res of the energy balance; positive Res indicates missing turbulent energy compared to the available energy.

Station		$Q_{H,EBC}$ [Wm^{-2}]	$Q_{E,EBC}$ [Wm^{-2}]	Res [Wm^{-2}]
A4	B	−11.1	4.1	90.0
A5	B	1.4	−6.9	91.1
A6	B	9.0	−9.7	74.8
A4	MAE	36.1	51.2	
A5	MAE	23.3	24.4	
A6	MAE	27.7	33.7	

Table 4. MAE and bias *B* between the optimised model runs and the unfilled, uncorrected observations during nighttime and the nighttime residual *Res* of the energy balance; negative *Res* indicates an excess of turbulent energy compared to the available energy, i.e. values around zero for the sum of turbulent energy fluxes compared to the (nearly always negative) available energy.

Station		Q_H [Wm^{-2}]	Q_E [Wm^{-2}]	*Res* [Wm^{-2}]
A4	*B*	−2.5	−13.7	−11.8
A5	*B*	−2.5	−1.9	−25.9
A6	*B*	−1.0	−10.7	−10.7
A4	MAE	5.3	15.6	
A5	MAE	6.2	6.4	
A6	MAE	4.2	12.2	

Table 5. Summary of the uncertainty estimations from various error sources. The flux measurement uncertainty is taken from Mauder et al. (2006), whereby type A corresponds to best rated sonic anemometer, e.g. Gill Solent HS, Campbell CSAT3, and omnidirectional probes, suited for general use flux measurements (type B) like Metek USA-1, following the recommendations by Foken and Oncley (1995). High and moderate data quality refer to flags 1-3 and 4-6, respectively, after Foken et al. (2004). The uncertainty of flux measurements is given as accuracy, the model input uncertainty as mean deviation from the reference run on a 95 % confidence level, and the other uncertainties correspond to mean absolute errors.

Error source	Q_H	Q_E
Flux measurements		
Type A instrument, high data quality	10 Wm^{-2}	20 Wm^{-2}
Type A instr., moderate data quality	20 Wm^{-2}	30 Wm^{-2}
Type B instr., high data quality	20 Wm^{-2}	30 Wm^{-2}
Type B instr., moderate data quality	30 Wm^{-2}	40 Wm^{-2}
EBC	range of the residual	
Gapfilling/EBC	32 Wm^{-2}	42 Wm^{-2}
Model input	20 Wm^{-2}	10 Wm^{-2}
Model parameter	24 Wm^{-2}	42 Wm^{-2}

Table 6. Flagging scheme for represenativeness within the upscaling concept: Detailed flags for target area, measurement quality and heterogeneity; $D_{j,\mathrm{mod}}$ is the difference between the modelled target flux and the modelled flux from land-use $j, j \in [2, 3, \ldots, M]$.
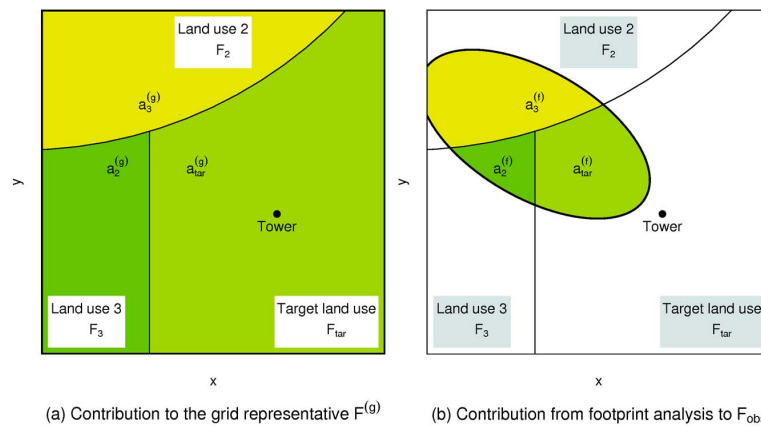
Flag name	Variable	Condition Q_H	Condition Q_E	Value	Description
Measurement flag	Accuracy (Table 5)	$<10\,\mathrm{Wm^{-2}}$	$<20\,\mathrm{Wm^{-2}}$	1	High instrument and data quality
		$<20\,\mathrm{Wm^{-2}}$	$<30\,\mathrm{Wm^{-2}}$	2	High instrument, moderate data quality and vice versa
		$<35\,\mathrm{Wm^{-2}}$	$<50\,\mathrm{Wm^{-2}}$	3	Moderate instrument and data quality, gapfilled values
Heterogeneity flag	$D_{j,\mathrm{mod}}$	$<20\,\mathrm{Wm^{-2}}$	$<30\,\mathrm{Wm^{-2}}$	1	Small differences between land-use types
		$<35\,\mathrm{Wm^{-2}}$	$<50\,\mathrm{Wm^{-2}}$	2	Moderate differences between land-use types
		$>35\,\mathrm{Wm^{-2}}$	$>50\,\mathrm{Wm^{-2}}$	3	Large differences, spatial integration strategy applied
Target flag	$a_{\mathrm{tar}}^{(g)}$		$1.0 > a_{\mathrm{tar}}^{(g)} > 0.8$	1	Predominant coverage by target land-use
			$0.8 > a_{\mathrm{tar}}^{(g)} > 0.5$	2	Target land-use builts up largest fraction in the grid cell
			$0.5 > a_{\mathrm{tar}}^{(g)} > 0.2$	3	Minor contribution of target land-use within the grid cell

Table 7. Overall flagging scheme for represenativeness within the upscaling concept, combining the target contribution within the grid cell (TFlag), the measurement quality (MFlag) and the heterogeneity between land-use types (HFlag).

TFlag	MFlag	HFlag	Error Q_H [$\mathrm{Wm^{-2}}$]	Error Q_E [$\mathrm{Wm^{-2}}$]	Value
1	1	1	20	30	1
1	1–2	1–3	20–30	30–40	2
2	1	1			
2	1–2	1–3	30–40	40–50	3
3	1–2	1			
1–2	3	2–3	35–45	50–60	4
3	1–3	2–3	>45	>60	5

(a) Contribution to the grid representative $F^{(g)}$ (b) Contribution from footprint analysis to F_{obs}

Fig. 1. Two different perspectives on a grid cell containing several land-use types, and the measurement tower situated on the target land-use. **(a)** Coloured areas contribute to a grid representative flux $F^{(g)}$, which is governed by the land-use specific fluxes F_j and their respective area $a^{(g)}$ within the entire grid cell. **(b)** Coloured shaded areas contribute to the observed flux F_{obs} at the tower, the respective area $a^{(f)}$ for each land-use specific flux is determined by footprint analysis. The location and extent of this source area is changing rapidly in time.

5217
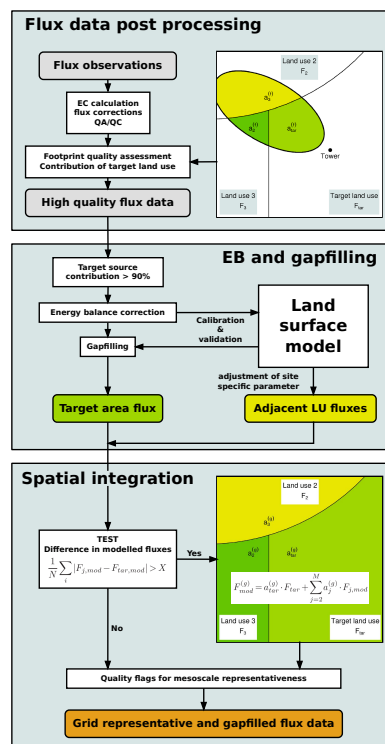
Fig. 2. Workflow for upscaling turbulent fluxes on a typical grid size of mesoscale models and satellite products with moderate resolution. Displayed is the target case, i.e. a substantial part of the measurements can be attributed to a single land-use type.
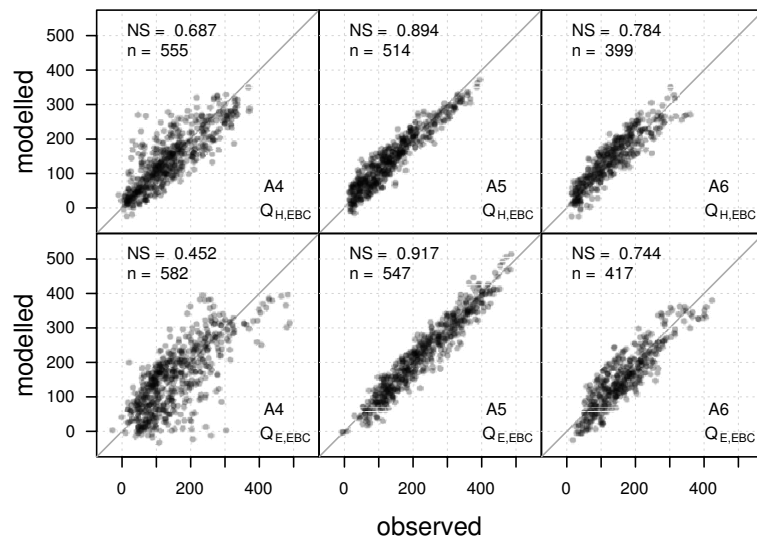
Fig. 3. Energy balance corrected fluxes vs. optimized model runs, subjected to gapfilling for each station; the upper panel shows the sensible heat flux, the lower panel the latent heat flux, with the stations A4 to A6 from left to right; n is the number of available observations ($n_{max} = 1392$) and NS is the Nash-Sutcliffe coefficient.
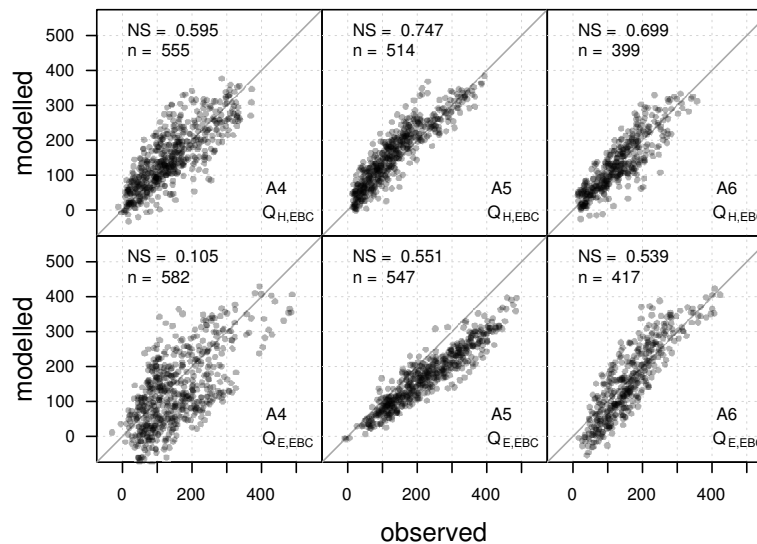
Fig. 4. Energy balance corrected fluxes vs. realistic model runs for each station; the upper panel shows the sensible heat flux, the lower panel the latent heat flux, with the stations A4 to A6 from left to right; n is the number of available observations ($n_{max} = 1392$) and NS is the Nash-Sutcliffe coefficient.
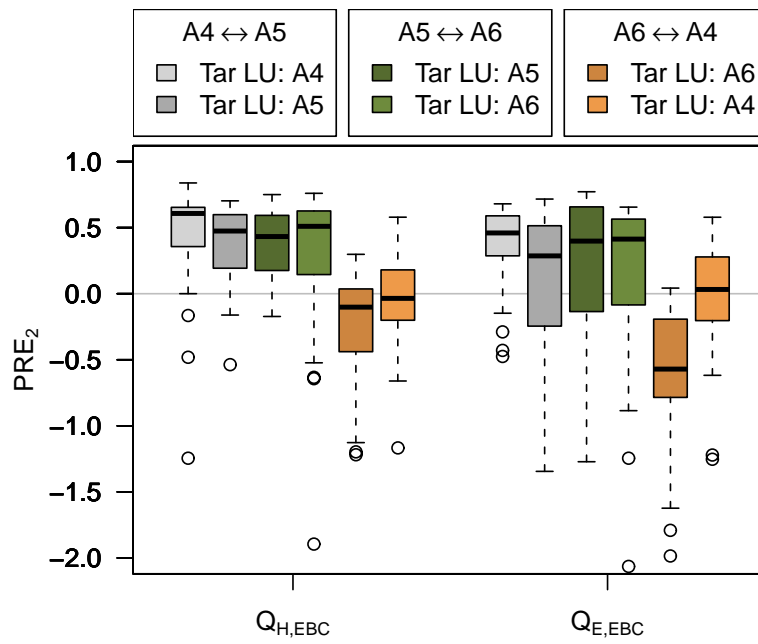
Fig. 5. Variation of daily evaluated PRE_2 for the energy balance corrected and gapfilled turbulent fluxes at all paired combinations of stations; the colour scheme denotes the specific combination of two stations, and for each combination the target land-use (Tar LU) is also changed.
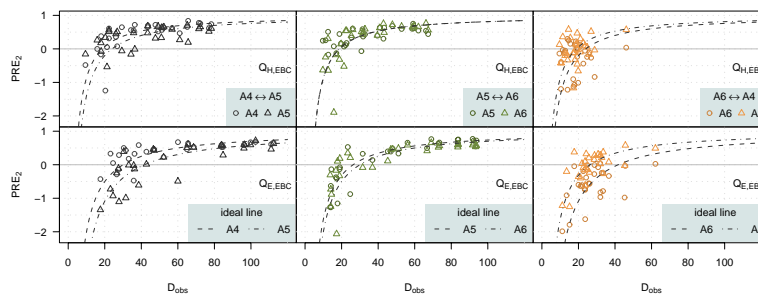
Fig. 6. Daily PRE_2 versus daily D_{obs}; the colours illustrate the pair of stations to be compared and the symbols (o) and $\triangle$ differentiate which of the two stations is assigned as target land-use.
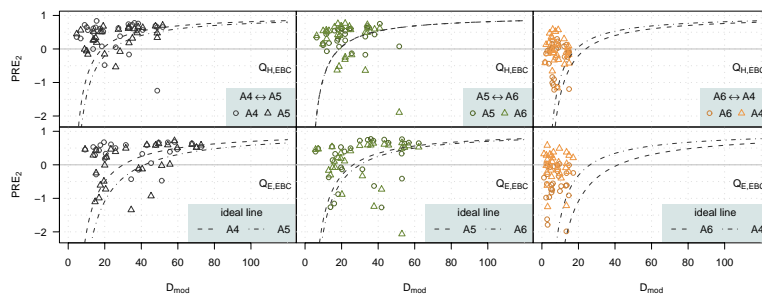
Fig. 7. Daily PRE_2 versus daily D_{mod} as a surrogate for D_{obs} (compare Fig. 6); the colours illustrate the pair of stations to be compared and the symbols (o) and Δ differentiate which of the two stations is assigned as target land-use.
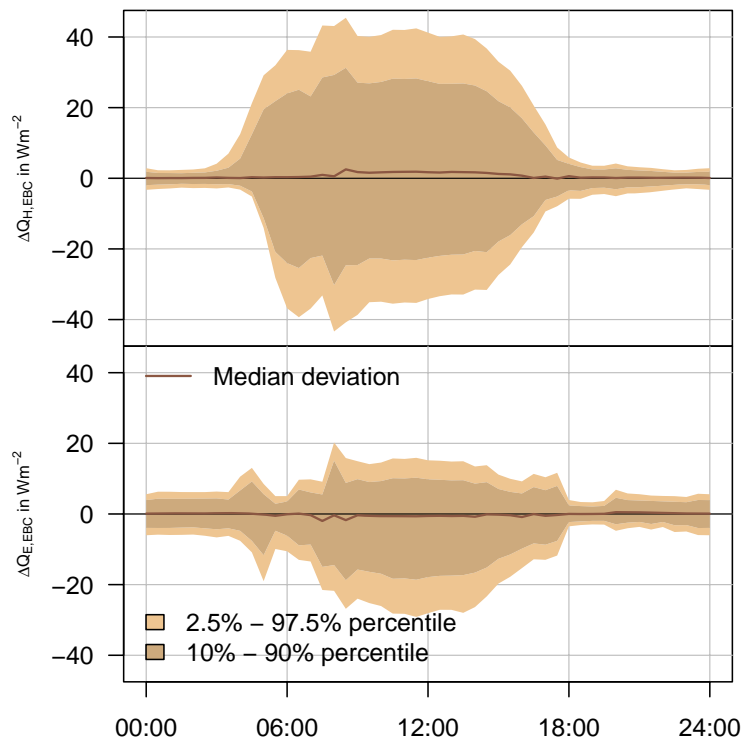
Fig. 8. Diurnal mean of differences between modelled fluxes from permuted forcing data and the reference modelled fluxes for A5 station; the red solid line is the median deviation and the coloured areas show the range of the percentiles containing 80 % (dark) and 95 % (light) of the differences.
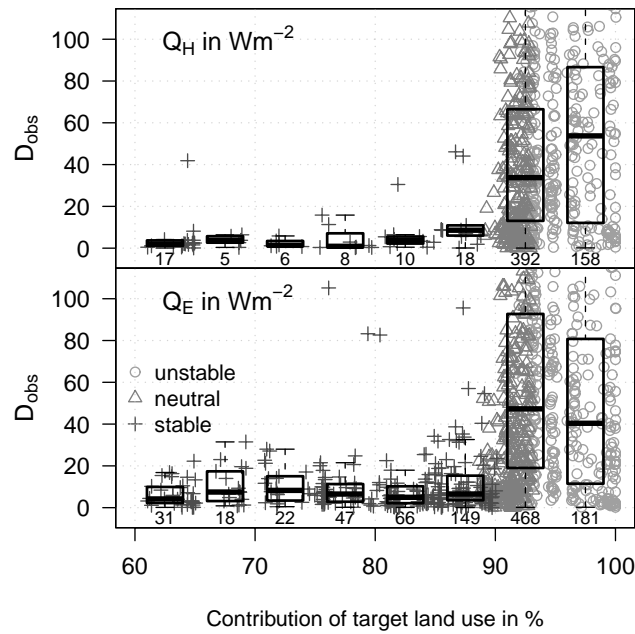
Fig. 9. Absolute difference between observations of A4 and A5 versus contribution of the target land-use according to footprint analysis in a virtual setup. Boxplots represent the data in classes of contribution (5 % width), with the number of data points within a class below the boxes. The symbols display the individual data points, with (+) for fluxes under stable conditions, (△) for neutral stratification and (o) for unstable conditions.