Hydrol. Earth Syst. Sci. Discuss., 8, C3251–C3263, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C3251/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "A framework to utilize turbulent flux measurements for mesoscale models and remote sensing applications" *by* W. Babel et al.

W. Babel et al.

wolfgang.babel@uni-bayreuth.de

Received and published: 5 August 2011

We thank the anonymous referee #1 for his general and detailed comments. These comments made us realize that our manuscript has some potential for misunderstanding in the modelling community. We start with a general reply in order to clarify the major misinterpretations. After that we give specific replies to the raised concerns and comments.

Scale considerations and aim of the paper

We believe that there is a misunderstanding regarding the aim of the manuscript and the applied methods. The overall aim is an upscaling of flux measurements with the eddy-covariance method on a pixel or grid size of remote sensing data and mesoscale models, which is recognized at least from referee #2. The upscaled flux, representative for one pixel (typically 1 km^2 when using MODIS data for e.g. land surface temperature), can be used to validate mesoscale flux simulations, driven by remote sensing data.

But the referees seem to underestimate the substantial gap in the scale of eddy-covariance data and the typical pixel or grid scale. The spatial scale of eddy-covariance data is highly variable depending on friction velocity and stratification which is for low measuring heights (up to 10 m) in the micro-β and micro-γ scale according to the classification by Orlanski (1975), while the pixel and grid scale is micro-α or even meso-γ. An overlapping in the stable case is possible (Foken and Leclerc, 2004), but in the unstable and neutral case the overlapping is small. All relevant references for the footprint problem are given in the manuscript, which are mainly Schmid (1997, 2002); Vesala et al. (2008). To overcome this problem of highly variable footprint areas in the paper a combination of footprint method and tile approach (Avissar and Pielke, 1989; Beyrich et al., 2006) was developed. From this point of view it should be clear that an upscaling on a 25 km^2 scale or catchments etc. cannot be the aim of the paper. For such a purpose mesoscale models, subgrid models or even Large-Eddy-Simulations must be applied.

The flux parameterisation in remote sensing or mesoscale models is based on the Monin-Obukhov similarity theory or a simplification like a bulk approach. Therefore data from two levels are necessary to determine the vertical gradient. The determination of the flux is a highly non-linear calculation due to the gradient approach and the influence of stratification. Therefore, it makes no sense to validate this approach against directly measured flux data when the input information (like surface temperature) is a mixed signal from a highly heterogeneous surface. This is the reason why we reduced the discussion of the upscaling mainly on two surfaces and introduced the multi-surface case only as theoretical approach. Furthermore we applied the method to an area without dramatic changes in the surface characteristics, otherwise the averaging of the surface parameters by remote sensing techniques is questionable due to the non-linear problem.

In summary, our aim is to provide high quality flux data for validation purposes of grid-based applications. The developed scheme combines the already existing features of QA/QC and gapfilling with considerations about heterogeneity and footprint and provides flux estimates including uncertainty ranges for each value (final flagging, Table 7). In the light of this target, the spatial integration via modelling of adjacent land-use types serves to estimate the uncertainty related to heterogeneity and it limits the error in case of substantial heterogeneity. We show, that it is possible to apply this feature, when a certain extent of heterogeneity is reached. But of course, it will be always preferable, if the measurements can be used directly. **We will point this out more clearly in the revised manuscript**.

General comments

"Theoretical basis: The authors suggest a pragmatic, straightforward method to derive the grid box mean flux without presenting any theoretical basis: The box is subdivided into patches of different land use. Fluxes for one land use class (i.e. target land) are obtained from measurements while modeling results are used for the other land use patches. Such an all-or-nothing scheme, putting full weight either on measurements or on modeling results is statistically suboptimal. Using model results to upscale and com- plete limited observations is a well developed field of research in meteorology named data assimilation. There, it is accepted that both sources of information should be used to obtain an optimal estimate by weighting the data according to the error covariance matrix – all-or-nothing schemes are not state of the art. Furthermore, the manuscript lacks any reference to data assimilation."

The topic presented in the manuscript is not related to any form of data assimilation and therefore no references to these studies are given. The main reason for this are missing permanent measurements of momentum, sensible and latent heat fluxes. The only one data basis is FLUXNET (Baldocchi et al., 2001) with 532 stations worldwide and e.g. 7 forest and 4 grassland stations in Germany. These stations were selected according to ecological reasons and the data are only non-realtime available, often only after several years. Furthermore usual meteorological models are not able to assimilate flux data directly. From this point of view we have not mentioned any assimilation concept or assimilation theory.

The underlying theory in this work is the tile approach for area averaging of flux data and the footprint method to determine the area from where the measured flux originates. We introduce both concepts (page 5167, lines 22 ff. and page 5168, lines 20 ff.) and provide the necessary details in Sects. 2.3 and 2.6 including all relevant references.

Remote sensing data and meso-scale models are not mentioned in the manuscript besides some references, because the aim of our study is the calculation of area averaged fluxes as ground truth for the validation of flux calculations from remote sensing data or mesoscale models (see Introduction, page 5169, lines 18 ff.). For example, a potential candidate may be the Surface Energy Balance System SEBS (Su, 2002). We agree with the referee that some changes in the title and an **extension of the introduction** could be helpful to avoid a misinterpretation of the aim of our manuscript. Therefore a possible **new title** can be "A framework to facilitate turbulent flux measurements as ground truth for the validation of mesoscale models and remote sensing applications".

"Adequacy of evaluation concept: Although the paper provides an extensive evaluation of the methodology based on LITFASS 2003 data, I question this proof of concept for the following reasons: i) The method is only tested for the situation of two land-use classes and the so called target case (Section 3.2.1). More land-use classes or the

mixed case (Section 3.2.2) are not considered. ii) The LITFASS 2003 campaign is limited to a short time period (about one month) with very special weather conditions (extremely dry summer of 2003). Thus is generalization of the findings is questionable. The authors even mention the need of significant tests (p. 5186, line 8), but unfortunately do not pursue this idea. iii) The flux differences between the considered land use classes are small."

We decided for this situation to test the method, because the ground truth is usually collected in widely homogeneous areas. As stated above, the non-linear model equations can only be validated over widely uniform terrain, where the remote sensing data does not contain a mixed signal of highly different land use types. From this point of view (and not only because it is more complicated) we reduced the number of different surfaces for evaluation within a grid cell to only two and selected sites which differ not extremely from each other. **We will explain the reason for our selection in more detail and add some rules for using ground truth to the manuscript**.

We agree with the referee, that any evaluation aiming at surface heterogeneity would benefit from having measurements from more land-use types over longer time periods. But as this is hard to get, we do not understand the criticism on the data set from the LITFASS-2003 experiment. The FLUXNET-stations mentioned above have only data sets over one land-use type. Experiments with flux measurements over different land-use types are rare and the given overviews show that data sets are only available for some weeks or month (Foken, 2008; Mengelkamp et al., 2006). Furthermore the specific weather situation is not relevant for testing the method. The benefit of the LITFASS-2003 data set is the good documentation; the data quality control and the installation of the flux stations make area averaging methods like the tile approach possible (Beyrich et al., 2006; Mauder et al., 2006). **We will add some remarks about the availability of flux data sets**.

"I was surprised to read the beginning of the third paragraph of the abstract: First it is claimed that application of the method to LITFASS 2003 shows the potential of the

approach. Then it is mentioned that the spatial integration would be rejected in this situation due to the small flux differences. I hardly dare to conclude that the benefit of the method is demonstrated for a situation which is not suitable for a practical application of methodology".

This conclusion is wrong from our point of view. The application of the method shows a clear benefit, even if the spatial integration is rejected, because the modelling of adjacent land-use types at least contributes to the uncertainty estimation given by the final flagging (Table 7) of the fluxes. The highly variable model performance on the daily basis originates from the change in surface conditions (Sect. 2.1) and enables us to judge, how large the difference should be, that the spatial integration can be applied (Sect. 4.3). The fact, that the spatial integration has to be rejected on the basis of the whole period, even if the outcome would be positive, is a matter of error consideration. It is a clear limitation for the spatial integration, but it does not compromise the scheme as a whole. **We will point this out more clearly in the manuscript**.

Specific comments

"Gapfilling: I would recommend to omit any gap filling for this study. Gapfilling is needed to calculate 'climatological' budgets, but not for any evaluation purposes – dates with missing values can easily be ignored. Gapfilling results in dangerous similarities between reference data and the data to be validated. This issue is shortly discussed on page 5192, but it can easily be avoided.":

Gap filling is a usual procedure for flux measurements (Falge et al., 2001a,b; Ruppert et al., 2006) because e.g. at night time for nearly negligible turbulent exchange the relevant measuring methods cannot be applied. During such conditions, typically the fluxes themselves and also the differences in fluxes between land-use types were small. The necessity to reject this existing data produces a large amount of gaps every day, and this problem is even increased by applying the energy balance closure (EBC) correction. Therefore gapfilling cannot be neglected in such a scheme, as it is a standard in the flux community. On the other hand, it is possible to resign gapfilling only for evaluation purposes, so we follow the suggestion. **We will abandon gapfilling in the evaluation procedure, but will retain it for the error considerations. This will change also Figs. 5–7**.

"1 km scale: At various locations in the text, it is mentioned that the method is designed for grid box with 1 km edge length. Please explain where does the scale '1 km' come from? Why is the technique not applicable for 500 m or 5 km grid boxes?":

This is related to the resolution of MODIS data in the thermal bands, where e.g. land surface temperature is retrieved from. See the replies above.

"Section 4.3 – evaluation of the threshold: In practical applications the critical point is how well D_obs can be inferred from D_mod. Please investigate this problem in detail. How good is the correlation between D_obs and D_mod?":

Comparing each single difference between observations for two land-use types with the differences between model simulations reveals no significant coherence. This is due to the highly variable model performance within the measuring period. We did not focus on this issue, as the important question was in our opinion the relationship of D_{mod} to the PRE_2. **We will add some details to this problem**.

"Balanced summary of the results: Both the discussion and the abstract are biased towards the positive results. Looking at table 2, it becomes obvious that the method is beneficial for two combinations of stations and shows neutral to negative impact for one combination. This two to one outcome needs be reported."

This is a misinterpretation from our point of view. In Sect. 4.2 the so-called "positive" outcomes were discussed (comparisons station A4 vs. A5 and A5 vs. A6, page 5187, lines 22 ff.) as well as the "negative" outcomes (comparisons station A4 vs. A6, page 5188, lines 3–6). As both A4 and A6 contain maize in a comparable growth stage, this outcome was highly expected and serves to demonstrate the variability within an

area which we would categorize as only one land-use type. If this variability would be comparable to the deviation between different land-use types, our categories of land-use types would turn out as not meaningful. Thus we considered this outcome simply as a control and did not emphasize it in the abstract, as it showed the expected (and necessary) result. **But we will add a sentence in Sect. 4.2 for clarification as our formulation seems to be misleading**.

"Focusing on upscaling: The paper should be better focused on its core topic: i.e. upscaling. For example, there is no need to discuss the instrumentation of the stations in detail or to explain the SEWAB equation (1)-(5). In addition, the paper will be easier to read if fewer references to other sections are used and certain topics are not separated into different sections: e.g. model evaluation techniques are both discussed in Section 3.1 and Section 4.1; discussion of threshold X (3.2.1 and 4.3). In the light of the quite simple theory, it might be possible to simplify the nomenclature that a list of symbols is no longer needed.":

We regret that our manuscript seems to be somewhat inconvenient for the reader. On the other hand, we think, that the instrumentation provides useful information, as it is directly related to the measurement uncertainty. Section 3.1 contains a more general view on the issue of representativeness, while in Sect. 4.1 this is condensed to our evaluated case. The threshold X occurs in Sects. 3.2.1 and 4.3 for the same reason. **We will revise the manuscript in order to eliminate unnecessary information and to reduce the number of crossreferences**

"Uncertainties and table 5: I wonder whether the deviations between model results and OBS derived in Section 4.4.1 (or Table 3 and 4) are an estimate for the model uncertainty – instead of the estimate at the end of page 5194. How do the values of 35Wm2 and 50Wm2 'follow from Table 5'? Why is there a need to list the instrument uncertainties in Table 5?"

The aim of this section is to discuss the various error sources introduced by the proposed routine in order to estimate an overall error of the upscaling scheme. Of course, the gapfilling uncertainty (EBC corrected observations vs. optimised model) is a model uncertainty like it is discussed in Sect. 4.4.2 (EBC corrected observations vs. realistic model). Nevertheless, we decided to handle them separately, because the contribution to the upscaling procedure were different for both types of model simulations and they were derived in a different way. **We will point this out in an additional sentence at the beginning of Sect. 4.4. Further this Sect. will be simplified, as the influence of gapfilling on the evaluation is obsolete**.

The values 35 $\mathrm{W\,m^{-2}}$ and 50 $\mathrm{W\,m^{-2}}$ simply exceed the largest individual uncertainty listed in Table 5. We decided to use these values in order to provide handy numbers on the one hand and to preserve the relationship of errors between sensible and latent heat flux.

The instrument uncertainty including quality checks for fulfillment of eddy-covariance prerequisites is an important error source. This topic has been already discussed (e.g. Foken and Oncley, 1995; Mauder et al., 2006, 2007). Nevertheless, it makes sense to list these results in some detail to illustrate the impact on the overall accuracy of the upscaling scheme, classified in Table 7: It is obvious, that the quality for representativeness strongly depends on the quality of measurements (MFlag: with an MFlag = 3 the overall flag cannot be better than 4). This is not only fixed to the instrument, but varies in time due to the dependence on the quality. Furthermore, the details for measurement uncertainty prepare for the classification of the MFlag fixed in Table 6.

"Need for calibrated SVAT: Is there really a need for a calibrated SVAT scheme? On page 5192, the MAE between OBS and uncalibrated SVAT is 39Wm2 and between OBS and calibrated SVAT 37Wm2. Is this small improvement of 2Wm2 worth the effort?"

As gapfilling will not be applied before evaluation, these values will be deleted from the manuscript, as they cause confusion: The optimised model runs perform substan-

tially better, which can be seen in Figs. 3 and 4. These values were not calculated for evaluation of model performance. Thus the observations without EBC correction were used for calculation in order to use a larger (and more representative) number of data to support the argument about gapfilling. Using the EBC corrected observations, the respective new values in page 5192 would be : mean MAE (EBC corrected observations vs. realistic model) = 47 $\mathrm{W\,m^{-2}}$, mean MAE (EBC corrected observations vs. optimised model) = 33 $\mathrm{W\,m^{-2}}$ and mean D_{mod} (optimised model vs. realistic model) = 44 $\mathrm{W\,m^{-2}}$. Please note, that the last value also changes due to an altered n, and the conclusion drawn in the manuscript remains the same.

"Title: The title is to vague, in particular the verb 'utilize'. It is misleading to mention 'mesoscale models' and 'remote sensing', because both applications are not presented in the paper. Suggestion: 'A method to upscale turbulent flux measurement using SVAT modeling'."

We agree, the title needs adjustment. Our suggestion is: "A framework to facilitate turbulent flux measurements as ground truth for the validation of mesoscale models and remote sensing applications". For explanation see "General comments" above.

"Quality flagging and table 6: I can not follow the reasoning to derive the rules of table 7. The underlying theory and assumption need to be outlined in more detail."

The quality flagging of flux data is a usual tool and in the last ten years world wide applied (Eigenmann et al., 2011; Foken and Wichura, 1996; Foken et al., 2004, 2011; Moncrieff, 2004). Thus it was consequent to add a new flagging system for the flux on the grid size. The restrictions of the error considerations were listed on page 5197, lines 9 ff., but the calculation is only outlined. **We will add more detail for a better understanding like in the following:** "The representativeness error has been calculated for all possible combinations of the MFlag, HFlag and the TFlag. Therefore the maximum error for each class of MFlag and HFlag were considered and weighted according to the minimum contribution of the target land-use to the grid cell, given by each class of the TFlag. In the special case of MFlag = 3 and HFlag = 1, i.e. the measurement error is larger than the error caused by heterogeneity, the measurement error is considered for the whole grid cell and the final flag is 4. (**This case will be added to Table 7**). All combinations were then sorted according to the overall error and categorized in five classes. The result is shown in Table 7."

References

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.

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.

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.

Eigenmann, R., Kalthoff, N., Foken, T., Dorninger, M., Kohler, M., Legain, D., Pigeon, G., Piguet, B., Schüttemeyer, D., and Traulle, O.: Surface energy balance and turbulence network during the Convective and Orographically-induced Precipitation Study (COPS), Q. J. Roy. Meteor. Soc., 137, 57–69, doi:10.1002/qj.704, 2011.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies

for defensible annual sums of net ecosystem exchange, Agr. Forest. Meteorol., 107, 43–69, 2001a.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Ta Lai, C., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., William Munger, J., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for long term energy flux data sets, Agr. Forest. Meteorol., 107, 71–77, 2001b.

Foken, T.: Micrometeorology, Springer, Berlin, Heidelberg, 308pp., 2008.

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

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

Foken, T. and Wichura, B.: Tools for quality assessment of surface-based flux measurements, Agr. Forest. Meteorol., 78, 83–105, 1996.

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., pp. 181–208, Kluwer, Dordrecht, 2004.

Foken, T., Leuning, R., Oncley, S., Mauder, M., and Aubinet, M.: Corrections and data quality, in: Eddy Covariance: A Practical Guide to Measurement and Data Analysis, edited by Aubinet, M., Vesala, T., and Papale, D., Springer, Berlin, Heidelberg, in print, 2011.

Mauder, M., Liebethal, C., Göckede, 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.

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: Intercomparison of eddy-covariance sensors and post-field data processing methods, Bound.-Lay. Meteorol., 123, 29–54, doi:10.1007/s10546-006-9139-4, 2007.

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.

Moncrieff, J.: Surface turbulent fluxes, in: Vegetation, water, humans and the climate. A new perspective on an interactive system, edited by Kabat, P., Claussen, M., Dirmeyer, P. A., Gash, J. H. C., de Guenni, L. B., Meybeck, H., Pielke Sr., R. A., Vörösmarty, C., Hutjes, R. W. A., and Lütkemeier, S., pp. 173–182, Springer, Berlin, Heidelberg, 2004.

Orlanski, I.: A rational subdivision of scales for atmospheric processes, B. Am. Meteorol. Soc., 56, 527–530, 1975.

Ruppert, J., Mauder, M., Thomas, C., and Lüers, J.: Innovative gap-filling strategy for annual sums of CO2 net ecosystem exchange, Agr. Forest. Meteorol., 138, 5–18, 2006.

Schmid, H. P.: Experimental design for flux measurements: matching scales of observations and fluxes, Agr. Forest. Meteorol., 87, 179–200, 1997.

Schmid, H. P.: Footprint modeling for vegetation atmosphere exchange studies: a review and perspective, Agr. Forest. Meteorol., 113, 159–183, 2002.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6, 85–100, doi:10.5194/hess-6-85-2002, 2002.

Vesala, T., Kljun, N., Rannik, U., Rinne, J., Sogachev, A., Markkanen, T., Sabelfeld, K., Foken, T., and Leclerc, M. Y.: Flux and concentration footprint modelling: State of the art, Environ. Pollut., 152, 653–666, 2008.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 5165, 2011.