General response

First of all we would like to thank the two reviewers for their helpful and inspiring comments. We added the publications of Wohlfahrt et al. (2010) and Charuchittipan et al. (2014) to the Discussion. The latter one was particularly interesting with regard to the proposed buoyancy flux ratio method. Moreover, we added in the Discussion a paragraph in which we discuss the possibility to use the PUB method in model inversion, Generalized Likelihood Uncertainty Estimation (GLUE) and Bayesian parameter estimation. In this context we also discuss in more detail the cons of the PUB method. In the following, the reviewer comments are given in Courier font and our reply is written in Calibri.

Reviewer #1

For the sake of completeness, it would be good to include the findings of the following two studies in the discussion section:

Wohlfahrt, Georg, Irschick, Christoph, Thalinger, Bettina, Hortnagl, Lukas, Obojes,Nikolaus, and Hammerle, Albin. Insights from Independent Evapotranspiration Estimates for Closing the Energy Balance: A Grassland Case Study. Vadose Zone Journal 9(4), 1025-1033. 2010.

Charuchittipan, Doojdao, Babel, Wolfgang, Mauder, Matthias, Leps, Jens Peter, and Foken, Thomas. Extension of the averaging time in eddy-covariance measurements and its effect on the energy balance closure. Boundary-Layer Meteorology 152, 303-327. 2014.

Indeed, these two publications are very interesting and relevant for our study. The following paragraph was added to the Introduction (blue letters indicate changed or added text passages):

„… Mauder and Foken (2006) evaluated EC flux data of the LITFASS-2003 experiment. The authors observed that the energy residual vanished almost completely if the flux averaging time was extended from 30 minutes (shortwave eddies) over 24 hours to 5 days (longwave eddies). The averaging time had a minor effect on the latent heat flux, but the sensible heat flux nearly doubled. Hence, in that data set, the energy gap could be mainly assigned to sensible heat. The approach to increase the averaging time for computing the covariance to 24 hours is questionable, because it appears that this procedure violates the fundamental assumption of stationarity. The authors argue that stationarity can be still assumed, because for the investigated 16-day time series the diurnal cycle was similar each day, and the trend of adjacent averages, which is the crucial stationarity criterion for the EC method, was smaller for 24-hour values than for 30-minute values. The finding that at some sites the energy residual may consist to a large extent of sensible heat was recently supported by an in-depth evaluation of additional EC flux data of the LITFASS-2003 experiment acquired over six different land use types (Charuchittipan et al., 2014)… “

To the Discussion we added these two paragraphs:

“…In the literature, a few studies compared Bowen ratio adjusted EC fluxes against a second independent method for measuring the latent heat flux. This provides some experimental hints on the robustness of the Bowen ratio method. Wohlfahrt et al. (2010) tested EC ET rates against independent estimates from micro-lysimeters at a temperate mountain grassland over two measurement campaigns. The authors come up with the recommendation to force the energy balance closure by adjusting for the average Bowen ratio, meaning that the energy balance is closed on a daily basis by dividing the measured half-hourly H and LE by the daily Bowen ratio. This implies that the Bowen ratio is conserved on a daily basis, but not necessarily the energy balance on half-

hourly basis. Scott et al. (2010) compared ET rates obtained with the EC method against the watershed balance over a period of five years in semi-desert grassland and desert scrubland catchments in the USA…"

"…Recently, Charuchittipan et al. (2014) proposed a further post-closure method. They suggest closing the energy balance based on the buoyancy flux ratio. In this approach, the fraction of the residual attributed to the sensible heat flux depends on the relative contribution of the sensible heat flux to the buoyancy flux. In general, this approach assigns larger fractions of the residual to the sensible heat flux than the Bowen ratio method does. In the context of the PUB, H fluxes calculated with the buoyancy flux ratio method would be in-between the Bowen ratio and H adjusted fluxes. The difference between Bowen ratio and buoyancy flux ratio adjusted fluxes depends strongly on the Bowen ratio. At very high Bowen ratios (>10) both methods result in very similar adjustments. At lower Bowen ratios, however, the difference between both methods increases. At a measured Bowen ratio of 0.2 and an EBC of 80%, for example, the Bowen ratio method would assign 17% of the residual to H, while based on the buoyancy flux ratio method this fraction increases to 86% (at 20°C), and the Bowen ratio shifts to 0.44. It remains to be seen whether this novel approach will prove its worth in future…"

P16914, l10: it should probably read "measurement errors" instead of "measuring errors"

Right. Corrected as suggested.

Foken (2008) is missing in the reference list.

The paper was added to the reference list.

P16928: In his short comment, Albrecht Neftel questions the validity of the data presented by Wolf and Laca (2007), and I agree with him that the paper shows "some surprising and counterintuitive results". The authors did not comment on the questions by Neftel and a final version of the paper has never been published in ACP. Moreover, scalar similarity works normally quite well in the high-frequency range. So, I would suggest to drop this reference as it is also not necessary for any further conclusions.

Thanks for this advise. We removed this part from the Discussion.

Reviewer #2

But this is also one of the greatest problems I have with the current manuscript that it does not provide formal uncertainties but rather a qualitative tool. Models that are outside the constructed uncertainty band are supposedly not so good. This gives me a right or wrong decision tool that is very coarse because I can only filter out the most obvious wrong model formulations. Can I use the uncertainty band in a model inversion? Are the proposed metrics, bound coverage and bound preference, suitable for that? Is a classical error measure such as chi-square possible with the method?

Thank you very much for this inspiring comment. Yes, we think that BC (band coverage) can be used in model inversion as objective function. Furthermore, BC could be used in the GLUE uncertainty analysis to distinguish between behavioral and un-behavioral model runs. And it could be probably used in Bayesian parameter estimation. We added the following paragraph:

"…In the present paper, PUB was not used to provide formal uncertainties, but as a qualitative tool to identify periods during which the model showed definitely structural deficiencies. This right-or-wrong

decision tool is quite coarse because it filters out only the most obvious failure periods. Beyond, it should be possible to use PUB, for example, in model inversion. Here, the BC could directly be used as objective function. One could either search in the parameter space for the set of parameters with the highest BC or search for sets of parameters above a prescribed BC threshold. In the latter case one would get a distribution of parameters. In the GLUE (Generalized Likelihood Uncertainty Estimation; Beven and Binley, 2014) approach, which is well established in hydrology, the PUB could be used as a criterion to distinguish between behavioral and un-behavioral model runs. Model parameterizations below a prescribed BC may be regarded as non-behavioral and are excluded from the further uncertainty analysis. In the frame of a Bayesian approach for parameter estimation (see e.g., Braakhekke et al., 2013) PUB could be used to constrain the likelihood function needed to compute the joined probability density…"

We do not advise to compute a chi-square statistics because for that one has to make the decision which post-closure method is the right one. This is not in line with the concept of the PUB approach.

I also found it confusing that the last 3.5 pages discuss the Bowen-ratio method and all other attempts in the literature to close the energy balance. It discusses basically the problems with all the other methods. But it does not discuss the new method. Where is the relation with the new method? Where is the discussion about the pros and cons of the new method compared to all the faulty old ones? It seems that the discussion about the former attempts might be more suitable for the introduction.

As a response on a comment of Reviewer #1, we removed the paragraph discussing the issue of scalar similarity (line 77-86 of the response letter). Moreover, we moved the paragraph about studies that are related to the H post closure method from the Discussion to the Introduction. We would like to keep the paragraph giving an overview of studies that investigated the robustness of the Bowen ratio method in the Discussion, because at the very end we relate the outcome of this review to our findings. We discuss cons of the PUB method now in more detail within the discussion about the possibility to use the method in model inversion etc. (see line 103-115 of the response letter).

tau of Eq. 5 is a strange measure. EBR will definitely not be normally distributed. So a histogram-based method is probably more appropriate.

We fully agree that the EBR is usually not normally distributed but this is not a prerequisite for the application of the approach that we propose here. Our intention to center the window on unity is to treat energy gaps and energy excesses in the same way. If one, for example, centered the window on the mean (in our case EBR=0.74) and took the window width that we used in our study then one would include datasets with an EBR larger than 0.44 and smaller than 1.04 in the analysis. This means, that one would accept energy gaps of up to 56% while in case of energy excess one would reject data with an excess of more than 4%. In our approach the center of the window is defined beforehand but the final setting of the width of the window depends on the distribution (histogram) of the EBR data. In that sense, this approach is histogram-based. Therefore, we would like to retain the use of tau as proposed.

I do not think that the term "post-closure method uncertainty band" is well chosen. If closure means the closing of the energy balance than this is no post-closing but rather a closing method. But the

word method is not fitting either; it is a validity band based on energy balance.

The term post-closure refers to the circumstance that the energy balance is closed after the measurement was performed, this is, ex post. Alternatively, one could try to close the energy balance ex ante on-site, for example, by measuring additionally flux and storage terms (see e.g. Jacobs et al., 2008 or Oncley et al., 2007). The reviewer is right, the uncertainty band itself is not a method, but with the wording "post-closure methods uncertainty band" we want to express that this band indicates the uncertainty related to the open question which post-closure method fits best at my site? To make the meaning of the term PUB more clear we modified the following sentence of the Material and Methods as follows:

"… The post-closure methods uncertainty band (PUB) is a proxy for the possible systematic error of EC flux data due to the unknown nature of the energy balance gap and the therefore open question which post-closure method fits best at the site under study. We define here that a PUB must fulfill basically two criteria:…"

Moreover, we use now "methods" instead of "method" to underline that the uncertainty band is related to the set of the three post-closure methods.

I could not figure out the origin of the error bars in the figures. So I could also not understand why some symbols had error bars and others not.

The measurement error is computed by TK3.1 (see p. 16917, line 11-13 of the Discussion paper). The reason that in some cases the error bars are not visible is that they are smaller than the size of the symbol. We added a phrase explaining this to the figure captions (Fig. 1 and Fig. 5-10):

"… The error bars indicate the random measurement error. In some cases, the error bars are smaller than the size of the symbol and therefore not visible…"

The lines, especially the dotted lines, are unreadable if the paper is printed in black and white. But they were very hard to distinguish also in the colour print.

We revised Fig. 2 and 4. In Fig. 2 we exchanged short dotted lines against short dashed lines and increased the line thickness. In Fig. 4 we print the lines in black (reference, tau=0) and in green (for different tau) (see below).

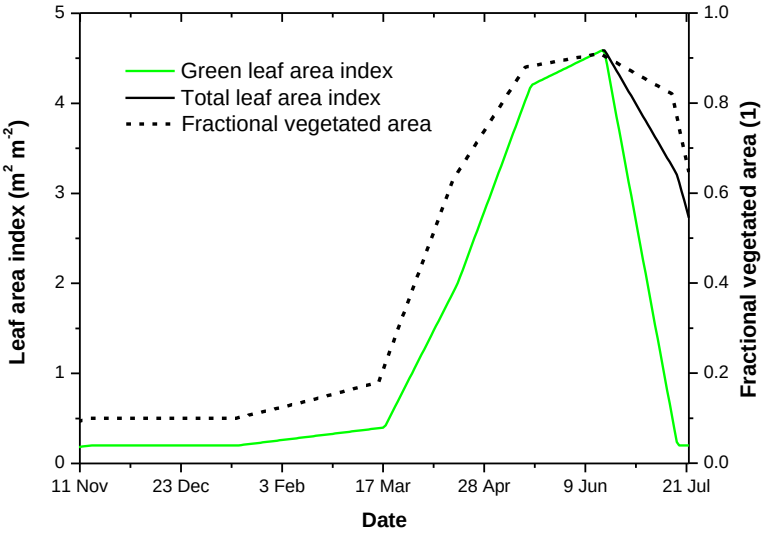

Figure 2

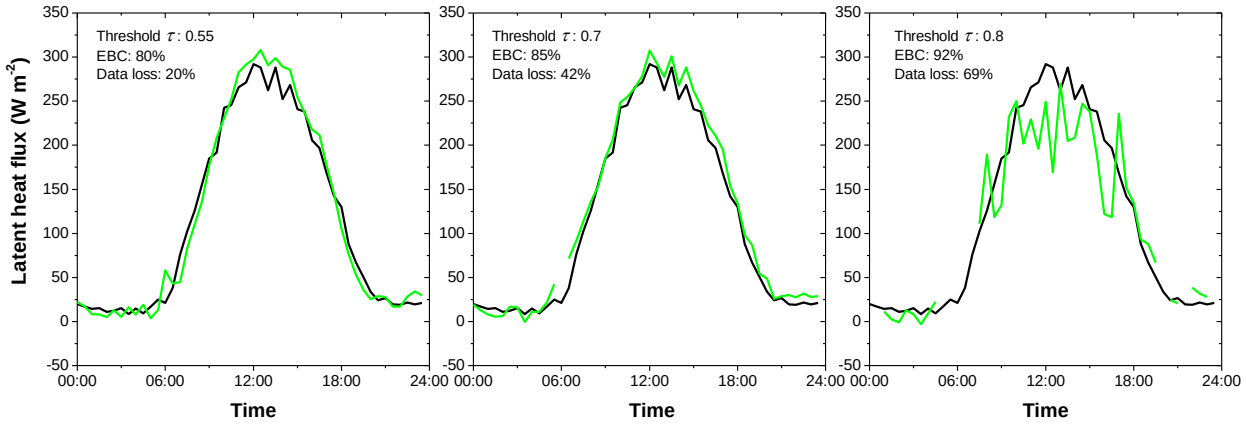

Figure 4

The new publication of Thomas Foken's group is missing, Charuchittipan et al. Bound-Layer Meteorolo 2014, which is the extension of Mauder and Foken (2006). It is also proposing another correction method.

Thanks for this advise. This is really a highly interesting and relevant publication for this paper. We added the publication of Charuchittipan et al. (2014) to the Introduction and Discussion (see line 38-41 and line 57-69 of this response letter). Among others, we compare for a hypothetical case the proposed buoyancy flux ratio method with the Bowen ratio method.