Interactive Discussion – Reply on comments of anonymous referee #2

We thank the reviewer for his/her constructive comments. He/she also provided a number of specific remarks that we will address in detail.

For clarity, we formatted the referees' original comments in blue italic, while our responses are formatted in black.

General comments:

> *This paper compares the evapotranspiration and the precipitation estimates by six lysimeters, on the one hand, and eddy covariance and tipping bucket rain gauge, on the other, for a grassland in Germany. The publication is of interest for instrumentalists as well as modelers looking for typical figures of uncertainty in those two components in this type of climate/vegetation conditions. However, for precipitation at least, the added value of the comparison is not shown since the reference set-up is rarely used when solid precipitation can't be neglected. For the evapotranspiration measurement, it's clear that the lysimeter performs well in this type of homogenous and simple vegetation cover, however for more complex crops the method can't be used for anything but evaluate a mere component, such as soil evaporation. This should be pointed out. The reference crop coefficient method should be substituted by a more complex energy balance equation in unstressed conditions, which leads to a modeled value that can be easily computed with the same amount of input data (plus some available parameters such as height and albedo, and allometric relationships well known for grass; did you measure LAI ?) as the crop coef method.*

We think that the general comments are handled by the responses given below. We are less negative about lysimeter measurements and think that these also give reliable estimates for more complex crops. We will consider the more complex energy balance method for calculating actual evapotranspiration.

Major comments:

1. *A weighing rain gauge with wind shield (such as the Geonor one) is usually recommended to measure solid and liquid precipitation, often in conjunction with snow pillows and snow height measurements. The underestimation of solid precipitation could be decreased by this system, it should be pointed out in the document. I guess using a combination of those instruments (which are easier to install than a lysimeter and have similar measurements footprints) could lead to a difference in total rainfall of the same order as that of the total evapotranspiration. Did you try classical wind correction algorithms for raingauge systems (even if you acknowledge that the error residual do not correlate well with wind)?*

Unfortunately a wind-shielded device for precipitation measurements was not available for our study. We will acknowledge in the discussion limitations of the current set-up. We compared uncorrected tipping bucket data to the lysimeters to investigate potential relationships of the found precipitation differences to weather conditions (e.g. wind, temperature) or precipitation aggregate. In the revised version of the manuscript we will also present a precipitation correction according to the method of Richter for some datasets during the summer period. On this basis we can expect an increase of the total yearly precipitation sum of approx. 5 % - 15 % compared to uncorrected measurements. We will complete these wind-corrected calculations and add monthly and yearly sums to the evaluation of precipitation.

2. *What is the difference between the 6 lysimeters with respect with the other components of the water balance (drainage, integrated soil moisture storage) ? (it could be useful to show cumulative differences between the 6 instruments and those 2 fluxes)*

We will add additional information regarding the water balance as suggested by the reviewer in the revised version.

Minor comments:

1. *P13808: error in relating eq. 6 and 7 and the methods to derive P and ETa (lines 4-6)*

Thank you. We will correct this.

2. *P13809L17 and L27: why 3h ? why 7 days ? Those 2 figures sounds fairly large to me, please justify; moisture status can change a lot in 7 days.*

We will add an explanation of the chosen time windows. Averaging the fluxes over a three hour interval guarantees a relatively small impact of sampling errors and therefore increases the reliability of the EBD calculation. On the contrary, the calculations would be strongly affected by non-stationary conditions for too large time intervals. Thus, the three hour time window is a compromise between the two sources of error. Further details are provided by Kessomkiat et al. (2013). This is similar for the EF estimations, which may be unreliable during single dark days with small turbulent fluxes. Kessomkiat et al. (2013) showed that a 7 day moving window reduces the instability of estimated EF compared to a 3 or 5 day moving windows. They suggested a minimum averaging period of four days for reliable EF estimates, because EF may show strong fluctuations with unmeaningful results and missing data for shorter time intervals.

3. *P13810L7: I don't understand how EBD3h(EF) is computed.*

EBD3h(EF) is the contribution of the latent heat flux (3 hour average) to the energy balance deficit according to the evaporative fraction for a certain measurement in time. We will clarify this in the revised version.

4. *P13818L23: why didn't you compute Eta with the full Combination Equation instead of the empirical Kc method ? (using actual roughness length derived from vegetation height for instance, esp. for such a well known grass cover)*

We consider this for the revised version of the manuscript, but still our data on grass height show gaps and insecurities, and we are not sure that this gives much better estimates. The empirical Kc method is an often used standard method.

Kessomkiat, W., Franssen, H.-J.H., Graf, A., Vereecken, H., 2013. Estimating random errors of eddy covariance data: An extended two-tower approach. Agricultural and Forest Meteorology, 171–172(0): 203-219.