Response to reviewer comments (AGRFORMET-D-14-00470)

We would like to thank the editor for extending the deadline what helped us to properly complete the revision of the manuscript. We thank also the reviewers for their helpful comments. Following the recommendations of both reviewers, we rewrote the discussion part and linked our results better with those described in literature. Following Reviewer #2, we searched for literature to correct our rainfall measurements. The World Meteorological Organization (WMO, 2009) performed a field intercomparison study with 25 different rainfall gauges. Among other, they tested the AGR100 gauge, which was used in the present study. For all tested gauges the WMO gives a formula to correct the rainfall gauge for catching, wetting and evaporation losses. We applied this correction to our rainfall data, which increased on average rainfall by 12%. Also, as recommended by Reviewer #2, we revised Table 2, and we give there now additionally the anomalies for the different observation periods.

During the revision process, moreover, we found a bug in the calculation of the matric potentials. The matric potentials sensors deliver a resistance which is converted in the data logger program into a matric potential. This conversion is a function of temperature. Accidentally, we used the top soil temperatures in this conversion. We recalculated the matric potentials and used for the depths 130 and 150 cm the mean annual temperature of the site (9°C). Based on the new matric potentials data and corrected rainfall measurements, we recalculated seepage and ET_{WB}. New results are now marked in blue. They generally follow the same pattern as the previous results, so that our main results and conclusions are unchanged. All other questions and comments we have addressed as described below.

Response to Reviewer #1

2.1 Main questions

1. The discussion section contains a lot of review material about related articles. It is good that the authors make the reader aware of the existing literature, but I think it would be good if the authors complement this by applying the general ideas to their own site. Barring experimental errors from the instrumentation, the two major schools of thought in EC can be summarized as storage versus heterogeneity effects (advection/secondary circulations). Which site-specific arguments would support one or the other, or both? With some six OPs available for two sites, can a distinction be made between local storage effects and heterogeneity effects and the relation with the EBR of the OPs?

We rewrote our discussion part and now we linked literature observations better to our results. In our view the energy residual is a combination of local storage effects and underestimated heat fluxes (e.g. secondary circulations). We discuss this in lines 461-511.

2. The paragraph of the error estimation for the soil water balance method is rather formal (e.g. that $\frac{\partial y}{\partial x}$ is the derivative of *y* with respect to *x* is self-explanatory in my regard and can be omitted) but it does not give a clear understanding of the practice: does *y* represent the bimodal Van Genuchten model, or ΔS, or something else? In my understanding there were 16 observations, so is $\sqrt{n}$ = 4 or does the number of observations refer to another quantity?
We revised this part, and the error estimation for ET_{WB} is described now in more detail and more concrete (*Lines: 329-337*).

Furthermore, how high was the variability between the 16 observations, i.e. how well is the footprint-weighted average of these 16 samples over a whole field site representative for the real average? Did the authors try a semi-variogram to check the spatial variability? What would be the uncertainty on the measurement related to the limited sampling? (e.g. right now the error associated to the WB method is of the same order of that of the EC method, would that still be the case?)
As suggested, we computed semi-variograms to check whether our assumption of uncorrelated sampling points is correct. The results showed that the 16 ΔS sampling points are not or only weakly spatially correlated. Computing the footprint-averaged ΔS with the geostatistical Kriging method instead of using simply the arithmetic mean of the 16 sampling points resulted in difference between 0.4 and 1.7 mm, what corresponds to a relative error below 0.5%. Therefore, at our study site using the arithmetic mean is fully justified. The standard errors of ΔS were quite small and ranged between ± 6-9 mm what corresponds to a relative error of 1.3 to 2%. We state these standard errors now in the Results section in subchapters 'Growing season 2012' and 'Growing season 2013'. "… Standard error of ΔS ranged in 2012 from 6 mm (1.3%) to 9 mm (2%) and in 2013 it was up to 8 mm (1.7%). The standard error of the water storage measured with SM1 sensors was on average 3 mm (1%)…".

3. Table 3: why has OP-4 the worst closure? Because it was earlier in growing season (the meteorological conditions seem fine) or is there a relation to its lower Bowen ratio? Can the authors connect this to the commonly cited causes in the literature?
The energy residual was in general higher at EC1 (40%) in comparison with EC3 (29%). We assume that the difference is induced by the heterogeneity of the surrounding. We discuss this point in the manuscript in lines 471-477. Most probably, the worst closure during OP-4 could be assigned to the additional heterogeneity caused by the differences in phenological development of the plants on different fields. OP-4 was early in growing season. In the Kraichgau region during this time some fields are already well covered with vegetation (e.g. winter cereals and winter rape) while others are still bare, prepared for late-covering crops, i.e. corn, potato, sugar beet (Imukova et al. 2015). Later in the growing season, fields are evenly covered with vegetation.

Imukova, K., Ingwersen, J. and Streck, T.: Determining the spatial and temporal dynamics of the green vegetation fraction of croplands using high-resolution RapidEye satellite images, Agric. For. Meterol., 206, 113-123, 2015.

2.2 Minor questions

4. 6789/20: Why do the orientations of the two anemometers differ?

The anemometers had the same orientation (170°). What slightly differed was the Licor-to-anemometer orientation. In the field an exact copy of the installation geometry from station to station is difficult to obtain. But an exact identical installation is also not really required, because the geometry of the sensor tandem is considered by TK3 in the flux calculation.

5. 6786/16: "Large eddies can be formed at the boundary of areas with different land use." This is slightly imprecise. I understand that their formation is linked to the presence of different land use types (and hence related to the presence of boundaries) but the (large) eddies would not be restricted to the boundary region. The authors declare that because the large eddies "do not touch the land surface, their transport of heat, water or gas is not detected by the EC station". Can the authors improve these sentences? One could argue that the large eddies do not directly touch the land surface, but this is unrelated to being undetectable by a single EC tower. (The reason for that would be that they're quasi-stationary.) Finally, the authors say that the large eddies are non-uniformly distributed: but distributed over what?

We rewrote the paragraph. The large eddy theory is now described more precisely. (Line: 82-101)

6. 6785/15: how much can it help (answer here in general)? Please be more specific. How much would it matter for the study sites? (the latter answer in the discussion)

We rewrote the paragraph in the Introduction and discussed it in the Discussion chapter. (Line: 52-57)

7. 6787/16: In this study (with its apparent focus on the latent heat flux) there would be no distinction with respect to the H post-closure. So I'm not convinced why it has to be mentioned at all. Was there an advantage of using the raw energy fluxes in those studies that has implications for the present study? If so, please clarify.

In the mentioned studies the authors were mostly interested in the patterns of fluxes and in these studies a closed energy balance was not important. We agree that this sentence is confusing. Therefore, it was omitted.

8. 6792/7: this was an average over the three classes or there were 3 averages, each for one stability class. How was the weighing done?

Described footprint model uses as input the covariance data processed with TK3 software (30 min resolution), particularly LE, H and momentum fluxes, as well as weather data (30 min resolution). Based on these data model estimates the footprint for three atmospheric stratifications (stable, neutral and unstable). The shown in the present manuscript footprint was calculated as an average footprint of these three classes, weighted based on the frequency of the different atmospheric stratifications. We added this information in the paper. (Line: 229 – 232)

9. 6800/15: how high was the correlation?
An R^2 of about 0.8 was reported in this study. We added the R^2 value to the manuscript. (Line: 435)

10. 6802/1 and 6803/11: make "most periods" more quantitative.
We corrected it. It reads now as: "…observation periods (OP 1 – 5)…" (Line: 509).

3 Technical corrections etc

3.1 References

11. In many cases the authors of the references are incorrectly specified in the manuscript, e.g. often an "et al" is missing, or only one author is named when there are in fact two authors. The authors should check their reference list and correct this. Simultaneously, the pronoun "he" is often used when it should be "they".
Corrected.

3.2 Formulas

12. Formula (3) is placed between two paragraphs without connection to the preceding sentence. The formulas can be read more fluently when they form (even when placed on a separate line) part of the sentences. (In that case also the punctuation has to follow standard spelling.)
We corrected it. Now the formula is better connected to the preceding sentence.

13. 6790/25: The EC ET units kgm−2 and mm = `m−2 are technically speaking not equivalent. The numerical values in these units will of course correspond when
_H2O = 103 kgm^{-3}.
We replaced kg m^{-2} with L m^{-2}.

14. Units are missing in equation (6).
We added the units of the heat of vaporization. (Line: 201)

15. 6795/23: I believe the notation θ_v was not introduced.
Right. We corrected it. (Line: 320)

16. In my opinion Gaussian error propagation is much more basic than the Akaike information criterion, so formula (11) is not needed. As mentioned in the main questions, the problem of this paragraph is not the theory, it's that the precise application is not clear (at least to me).
We added a more detailed description of the error estimation and changed Formula 11.

3.3 Structure and content

17. (An outline of the paper is missing at the end of the introduction, but the structure of the paper is also clear enough without.)
We think that the structure of the paper is quite straight forward. Therefore, we did not add an outline to the manuscript.

18. In Table 2 the data should be added because OP-0 appears in Figure 7.
We added corresponding data for the OP-0 in the Table 2.

19. 6794/11: Perhaps add the values for the Akaike information criterion of the tested models to highlight the selection of the preferred model. (As of now a number of models is simply listed but only one is chosen, so I don't see why naming all the other models is relevant. In my opinion this could be omitted.)
We removed the list of models and rewrote this part (Line: 286-288)

3.4 Other

20. 6791/10: it's of course completely equivalent here, but why not use as unit Kelvin instead of degrees Celsius
Because in the above introduced empirical equation of Foken (2008b) for computing the heat of vaporization the temperature is given in °C, we prefer to keep also here the unit in °C.

21. 6792/10: I believe the installation height was mentioned in 2.2.1 and was 13 cm higher there.
For the calculation of the footprint we used flux and weather date of the year 2010. In 2010, the installation height of the turbulent complex was indeed somewhat lower than in 2012. This difference, however, has only a very little effect on the footprint calculation.

3.5 Language suggestions

22. 6784/4: I prefer "energy imbalance" or "energy budget non-closure" instead of energy gap, the latter reminds of energy supply and solid state physics.
We rewrote abstract therefore it is no longer an issue.

23. 7: please write both H and LE slanted (or all upright) throughout the manuscript
Corrected.

24. 23: "vegetation period" could be replaced by the more standard "growing season"
We replaced everywhere "vegetation period" with "growing season".

25. 24: **a verb is missing**
We corrected it. It reads now as: "*Our results indicate that the energy balance gap is made up by other energy fluxes and unconsidered or biased energy storage terms."* (L: 35)

26. 6785/9: the ground heat flux
Corrected. (L: 52)

27. 21: in this context, I would prefer the term "energy balance ratio" EBR instead of the energy balance closure EBC
We did not want to mix different terms, in order to be clear for the reader, and therefore we have chosen, in this case, an "energy balance closure" term.

28. 6786/3: during the night when fluxes are low

We corrected it. (L: 70)

29. 14: concluded instead of discussed
Corrected as suggested. (L: 81)

30. 17: analyzed instead of performed
We rewrote this part.

31. 19: quantified instead of measured
We rewrote this part.

32. 6787/9: in this context I would prefer “residual” instead of gap
We replaced “gap” with “residual”. (L: 112)

33. 6788/3: in the region Kraichgau
Corrected. (L: 125)

34. 8: and the landscape consists of gently sloping hills
Corrected as suggested. (L: 130)

35. 18: surrounded by other agricultural fields
We corrected it. (L: 139)

36. 6789/15: the fluctuations of the two quantities
Corrected. (L: 159)

37. 20: the Licor, the CSAT3
From our point of view, in this case, “the Licor-CSAT3 separation” is clearer. Therefore we would like to keep it like this.

38. 6791/6: perhaps put brackets around “using the mean of [. . .] the plates” to make the sentence more clear
We prefer to keep it as it was. We think the formulation is clear.

39. 6792/13: Matrix/matric potential can both be used, but in other uses it is one **matrix** and two **matrices**. Occurs elsewhere in the text as well.
Corrected.

40. 6794/15: of the two
We corrected it. (L: 289)

41. 6796/1: The latter two periods. (the last could also mean OP-5 and OP-6)
Corrected. (L: 326)

42. 6797/17: during instead of over
Corrected as suggested. (L: 363)

Response to Reviewer #2

General

The energy balance closure of winter wheat stands is analyzed. Different methods to close the energy balance gap are compared: the Bowen ratio method, assigning the complete gap to the latent heat flux, or assigning the complete gap to the sensible heat flux. A comparison is made with a water balance method, which is taken as truth. It was found that assigning the complete gap to the sensible heat flux gives in general the best results.

My main concern is the quality of the water balance method given the uncertainty with respect to the gradient of the hydraulic potential and the hydraulic conductivity function.

We agree that the drainage and capillary rise are the two fluxes with the highest relative error, and we are of course aware about the spatial variability of the hydraulic conductivity at the field scale. The WB method is most accurate during periods of low rainfall and consequently negligible drainage rates (Schume et al., 2005; Wilson et al., 2001). These conditions were well fulfilled during OP 4 and 5. During OP4 and OP5 we have a nearly perfect match between the WB method and the H post-closed EC ET data. The results that we obtained during OPs with higher seepage flux (OP1-3) are in line with the findings of OP4 and 5. Therefore, we are confident that the determined seepage fluxes are in the right order of magnitude and that the total error, which is relatively low due the small absolute flux, is in an acceptable range. We added a paragraph in the Discussion discussing this issue (line 414-423).

I believe we have a large uncertainty and the uncertainty of ET with this method is very high. A further aspect is that precipitation is uncertain as well, and might often be underestimated with standard methods. Therefore, ET would be underestimated as well with water balance methods. I think a more thorough assessment of the uncertainty of the water balance method, insights and discussion of its limitations are needed for the paper.

As mentioned above, after thorough check of literature, we agree with Reviewer#2 that our rainfall measurements need indeed a correction. Tipping buckets rain gauges and rain gauges installed above the surface in general systematically underestimate rainfall. For this purpose we used a correction scheme for our rain gauge ARG100 which was obtained by The World Meteorological Organization (WMO) in the frame of an intercomparison study of 25 rain gauges under field conditions (see Reference list: WMO, 2009, p. 57)). The applied correction increased rainfall on average by about 12% in both years. Similar underestimation rates were also observed at the Rothamsted weather station (www.era.rothamsted.ac.uk). They compared over an 8 year period the ARG100 with a rain gauge installed within a turf-wall enclosure that means at the level of the land surface. They found that on average the ARG100 underestimates rainfall by 10%. We added a short description of the WMO correction to the MM (Line 172-174) and mention in the Result section that the correction yielded on average 12% higher total rainfall (Line 362).

My second main point is that it is unclear what we learn from this paper. The authors have a nice summary of all results obtained with respect to the energy balance gap in the discussion. Sometimes the best performance is obtained with energy balance closure using the Bowen ratio method, sometimes using a H closure, and sometimes a LE closure. In this case it could be the LE closure. It is only a result for a single site among many sites. Additional motivation is needed for the publication of these results.

We added a paragraph to the Discussion in which we give additional motivation and explain why it was for us important to perform such a site-specific study. Please see lines 461-469.

Detailed comments

P6784, L24-L25: Rephrase.
Thank you for the comment, the verb was missing. It reads now as: *"Our results indicate that the energy balance gap is made up by other energy fluxes and unconsidered or biased energy storage terms."*

P6785, L11-L12: In other works those storages were included, and later you argue that they could play a major role. I would therefore reformulate this sentence.
We rewrote this paragraph. (Line: 50-57)

P6789, L5: Repetition.
We corrected it. (Line: 149).

P6790, L7-L10: Given the usual strong spatial variability of soil hydraulic parameters, can fluxes be reliably estimated?
Please see above and lines 414-423 in the manuscript.

P6792, L14: "matrices" instead of "matrixes".
We corrected it.

P6793, L16-L27: This comment relates to the comment before. Both the gradient of H with respect to z and the function K(h) are very uncertain. From the measurements, and given measurement errors and spatial variability it seems to be difficult to precisely estimate deltaH/deltaz. Concerning the estimation of K, it can be expected that lab estimates of K and field values differentiate substantially, given the very different scales.
Please see above.

P6795, L4-L6: Is there a bias in the estimated soil moisture content as machine tracks have a very different soil moisture content compared to the areas between the machine tracks?
Measurements of the water content in machine tracks were not performed. At a working width of 15 meters and a track width of 0.3 m, about 2% of the field area are machine tracks. This small fraction of machine tracks is too small to cause a significant bias.

P6796, L17-L18: How is this determined? Which method? What does this exactly mean?
The random error of the fluxes consists of the instrumental noise error of the EC station and the stochastic (sampling) error. The method was mentioned in the Chapter 2.2.1 (Line 195-197). The random error was estimated with TK3 software where automatic quality assessment algorithm developed by Mauder et al., 2013 is implemented. This algorithm is applied as part of the raw data processing and so assesses the data quality based on direct analysis of high-frequency data.

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H. P., Schmidt, M. and Steinbrecher, R.: A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, Agric. For. Meterol., 169, 122-135, 2013.

P6797, L11-L14: the order of figures should be changed here: Figure 5 is introduced before Figure 4. Before Fig. 8 was already introduced, so this does not seem to be very logical.
We corrected it.

P6798, L10-L12: Why?
We rewrote this sentence. (Line: 387)

P6799, L8-L25: This complete text block was not analyzed in this paper. It fits in the introduction, but should not be repeated in the discussion, or only very shortly, indicating that physical mechanisms on the energy balance gap were not analyzed in further detail in this paper.
We shortened this text block.

P6800, L13-L14. A similar conclusion was reached by Alexander Graf et al. in a recent paper in WRR. However, a possible explanation could also be the underestimation of precipitation by tipping bucket measurements.
We discussed the uncertainty of rainfall measurements above and corrected the rainfall data for catching, wetting and evaporation losses.

P6801, L3-L4. Another study was by Gebler et al. (2015, HESS), where they found that adjusting ET from EC by the Bowen ratio method gave the best fit with ET measured by lysimeter.
We added this reference. (Line: 449-451)

P6801, L27-L28. It is not so clear then, if ET is already measured, why we still would want to install an EC-tower.
We expanded the sentence and now it reads as: "*Therefore, it is advisable in case of long term experiments to perform for each site an independent measurement of LE to identify the most suitable post-closure method".* (Line: 463)

P6802, L19. Rephrase.
We corrected it.

P6803. Conclusions. The conclusions cannot be read independently from the rest of the paper. Introduce some details about the sites (e.g., where) and re-introduce again the abbreviations.
We mentioned now that study site locates in southwest Germany and re-introduced the abbreviations.

Table 2. This table is not very informative like this. Think how to present these data in a better way. A possibility is to present anomalies. You could use the unit per day.
We corrected it.