Reply to Referee #3 (hess-2016-252)

We thank Referee #3 for the valuable comments (in italic in the following) and will address them point-by-point in the following.

Hirschi et al. reports an intercomparison of eddy covariance (EC), lysimeter, and catchment observations of evapotranspiration (ET) from a small catchment in Switzerland. The description of the methodology and intercomparison is very thorough, and the results add a new ecosystem to this type of intercomparison (many previous lysimeter- EC studies have been on irrigated agroecosystems in semi-arid or arid regions). The study is generally well-written and presented. There is, perhaps, a bit too much emphasis on the EC error terms and too little discussion of the underlying site meteorology and vegetation of the lysimeter, EC footprint, and catchment as a whole. How does vegetation density and greenness in the EC footprint compare to the lysimeter and the overall catchment? How significant is advection in this site/region?
We are happy to see that the presented inter-comparisons of the parallel, multi-year measurement records is valued by the reviewer. We extended the site description by adding an aerial map of the site and by also discussing a previous footprint analysis (see also reply to Reviewer #2).

In addition, we investigated the amount of vertical advection of latent and sensible heat (see e.g., Paw U et al., 2000; Casso-Torralba et al., 2008) using the notation $F_v = \overline{w}_z(\overline{m}_z - \langle m \rangle)$ by Lee (1998), where $\overline{w}_z$ is the mean vertical wind velocity at height z, $\overline{m}_z$ is the measurement of moisture or temperature at height z, and $\langle m \rangle$ denotes the vertical average of the moisture or temperature measurements up to the height z. It is thus expressed as the vertical gradient of moisture or temperature (defined as the difference between the vertical average and a specific level) multiplied by the mean vertical wind speed at a specific level. F_v is scaled to W m^{-2} using density of air, latent heat of vaporization or specific heat capacity of air. Tilt correction of the sonic anemometer for obtaining mean w was done applying the planar-fit-method (as mentioned in the manuscript). Note that since the 2-m level is the lowest, we could only infer an estimation of vertical advection based on the temperature gradient between the 2 and 5 m levels and the moisture gradient between the 2 m and 9 m levels. The results reveal that the magnitude of vertical advection of latent and sensible heat is small (on average at most around -0.1 W m^{-2} respectively 0.05 W m^{-2} at noon, see Figure 1 below) compared to the respective average turbulent fluxes (less than 1%). We thus assume that vertical advection plays a negligible role in the energy balance calculations.

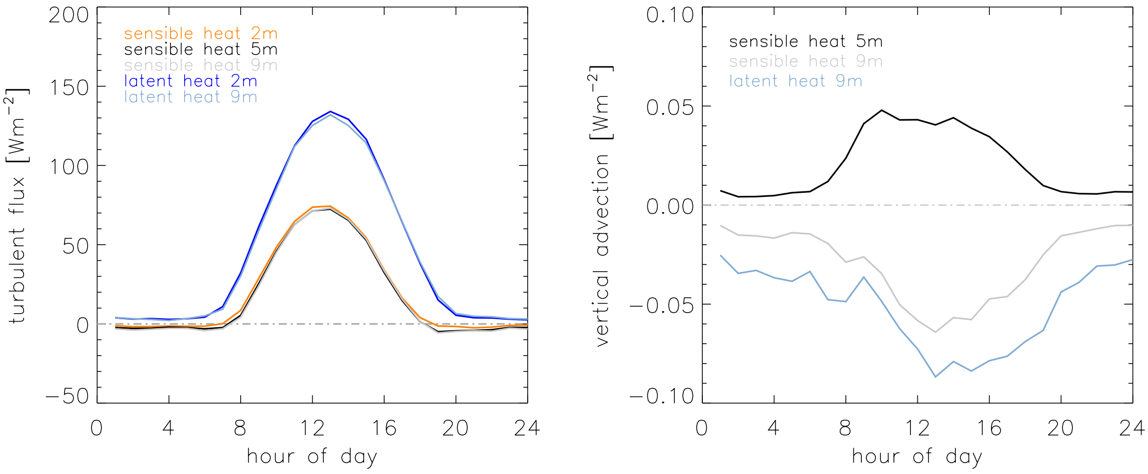

Figure 1: (left) Mean daily cycles of the turbulent fluxes at the different measurement levels and (right) vertical advection at these levels.

For horizontal advection, measurements for a quantitative estimate are not available at the site and its surroundings. Possible reasons for horizontal advection include slope drainage and heterogeneous land cover. Concerning the first reason (slope drainage), the wind from the direction of the closer south-facing valley slope is masked in the whole analyses as it includes the tower (see Section 2.3). Concerning the impact of heterogeneous land cover, we estimated the possible effect on energy balance closure by separating the closure analyses into three wind sectors (i.e., east, west and south wind directions; note that sector north is completely masked due to the presence of the tower). While the sector west (i.e., the main wind direction) has homogeneous land cover and horizontal advection should not be relevant, the sector east is potentially impacted by a small street and a farmhouse (see Figure 2 of the manuscript). Note that we focus on daytime here in order to rule out biasing due to the differing distribution of nighttime fluxes among the wind sectors. The results of these analyses reveal that the energy balance closure is rather independent of the wind direction (see Figure 2 below). The slope and R^2 of the regression analyses are similar for all three wind sectors. This also holds for the daytime ratio of the total amount of the turbulent heat fluxes to available energy, which amounts to 86.5%, 86.8% and 82.1% respectively for the east, west and south sectors. This robustness in the closure independent of the wind direction indicates that horizontal advection is not of great importance at the site.

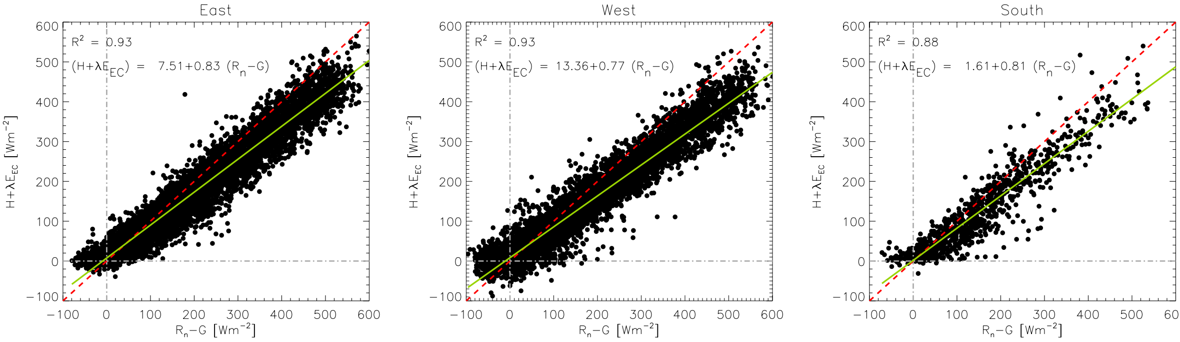

Figure 2: Energy balance closure as evaluated from the ordinary least squares regression between the daytime hourly estimates of the turbulent fluxes against the available energy, separated by the three 90°-wind sectors. (left) Regression based on hours with wind from sector east, (middle) from sector west and (right) from sector south. The dashed red line within each panel indicates the 1:1 line, the green line the regression line.

These analyses are now introduced in Section 2.3 and the corresponding results are presented in Section 3.2, with the Figures 1 and 2 being added to the supplementary material.

Do you have larger differences between the EC and lysimeter under specific meteorological conditions (certain wind direction/speeds, time of day, vapor pressure deficit)? Including some of these comparisons will help contextualize your results against other comparisons such as BEAREX reported by Alfieri et al. and the earlier study by Ding et al. – see doi:10.1016/j.agwat.2010.08.001).

We expanded the lysimeter vs. EC comparisons (focusing on E_{EC_BOWEN} only) to include specific meteorological conditions. In particular, we had a closer look at the statistics R^2, RMSD and relative bias (i.e., $\frac{E_{EC\_BOWEN}-E_{LO}}{E_{LO}}$) for the three wind sectors (see above) vs. different wind speeds, for the wind sectors vs. different times of the day (all-day, daytime and nighttime), and for high and low vapor pressure deficits vs. different times of the day. The results are presented in Section 3.3 and three supplementary tables (see Tables 1-3 below).

Table 1: Hourly data based statistics (R^2, RMSD, relative bias, number of data) comparing EC (E_{EC_BOWEN}) with lysimeter based evapotranspiration (E_{L0}) for different wind directions (i.e., sectors east, west, south and all together) vs. high and low wind speeds (i.e., greater and lower-equal than the median wind speed). Statistics are based on measured values only (i.e., excluding gap-filled data) and masked for precipitation. Units of RMSD in mm, R^2 and relative bias as fractions.

	high wind speed				**low wind speed**			
	R^2	RMSD	Rel. Bias	Number	R^2	RMSD	Rel. Bias	Number
East	0.79	0.090	-0.024	3707	0.81	0.067	-0.071	4777
West	0.77	0.081	0.041	9435	0.71	0.040	-0.271	10176
South	0.76	0.017	0.138	132	0.69	0.077	0.187	1004
All	0.78	0.084	0.016	13274	0.79	0.052	-0.112	15957

Table 2: As Table 1, but for different times of the day (i.e., all-day, daytime and nighttime) vs. high and low water vapor pressure deficits (i.e., greater and lower-equal than the median vapor pressure deficit).

	high vapor pressure deficit				**low vapor pressure deficit**			
	R^2	RMSD	Rel. Bias	Number	R^2	RMSD	Rel. Bias	Number
All	0.76	0.088	-0.023	14967	0.28	0.037	-0.086	14968
Day	0.67	0.107	-0.037	8043	0.53	0.063	0.147	8044
Night	0.06	0.031	-0.247	6922	0.00	0.027	-0.568	6925

Table 3: As Table 1, but for different times of the day (i.e., all-day, daytime and nighttime) vs. wind from different directions and without RMSD.

	All			**East**			**West**			**South**		
	R^2	Rel. Bias	Number	R^2	Rel. Bias	Number	R^2	Rel. Bias	Number	R^2	Rel. Bias	Number
All	0.80	-0.029	30002	0.81	-0.045	8602	0.78	-0.031	20242	0.71	0.175	1158
Day	0.75	-0.001	16154	0.77	-0.035	6495	0.74	0.015	8735	0.68	0.192	924
Night	0.04	-0.374	13847	0.03	-0.444	2107	0.04	-0.359	11506	0.00	-0.573	234

The agreement between E_{EC_BOWEN} and E_{L0} (visible in R^2 and the relative bias) is worst during nighttime when evapotranspiration is low and less variable (Table 2 and 3). For the same reason, also low vapor pressure deficit worsens the statistics, as evapotranspiration is also lower in such conditions. Moreover, the statistics during southern wind directions are worse than for the other wind sectors (however, based on much less data). Wind speed on the other hand does not seem to have a strong impact on the agreement between EC and lysimeter evapotranspiration, except for the increase in relative bias for low wind speed during western wind directions (Table 1).

Some other comments: 1. I agree with Reviewer 2 that more results about G need to be reported. I realize that much of the meteorological data may have been reported in earlier studies, but I think you need to at least refer back to these data.

We added an analysis of the heterogeneity of surface and 5-cm soil heat flux and of the soil temperature measurements in the supplementary material (see Figure 3 below).

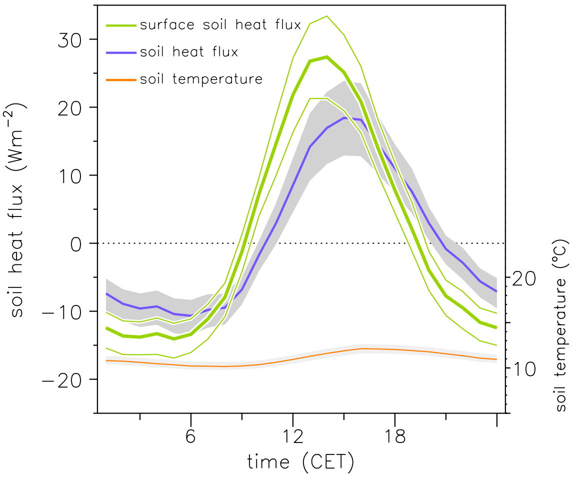

Figure 3: Daily cycles of surface (green lines) and 5-cm (purple) soil heat fluxes, as well as of soil temperature (5 cm depth, orange). For the latter two, solid lines present the data based on the averages of the three heat-flux plates and the three soil temperature sensors respectively, while the range is based on the data from the three individual sensor locations (and displays the minimum and maximum values respectively). For the surface soil heat flux, the estimate calculated from the averaged heat-flux plates and temperature sensors is displayed (thick line), along with a minimum and maximum estimate based on the individual sensor locations (thin lines).

The effect of the correction based on Fuchs and Tanner (1968) is clearly visible and leads to a shift of the daily cycle of the surface soil heat flux vs. the 5-cm soil heat flux, and to an enhancement of the daily amplitude. The range of the soil temperatures amounts to 0.9 °C on the average, while surface soil heat fluxes show a mean range of 6.7 W m^{-2}. Especially during nighttime, this amount is substantial compared to the available energy of around -25 W m^{-2}. These results illustrate the spatial heterogeneity of the surface soil heat flux footprint and underline the importance of employing a set of several soil heat flux sensors in order to obtain spatial representativeness of the data. This is now mentioned in Section 3.2.

2. With respect to energy budget closure for the EC tower, you assume that canopy energy storage and energy storage due to photosynthesis are negligible. I do not share this assumption. Photosynthesis storage can be significant in productive grasslands and is not corrected by diurnal averaging. Canopy energy storage can be averaged out with daily energy balances. Correcting fluxes by energy balance on a daily basis can also significantly improve energy budget closure as reported by Leuning et al. (2012) and Anderson and Wang (2014 – see doi:10.1016/j.agrformet.2013.09.012). How do your EC ET calculations change if you use daily closure of energy fluxes instead of hourly closure?

We tested this effect by comparing hourly vs. daily energy balance closure. Figure 4 shows this comparison based on days where maximal five of the hourly values were gapfilled, which leaves 462 days of valid EC observations. As mentioned by the reviewer, the energy budget closure slightly improves on daily time scales: regression slopes increase from 0.76 to 0.84, and R^2 from 0.95 to 0.97. This indicates a potential effect of diurnal storage variations on the energy balance closure. Possible storage effects and their averaging out by daily energy balance analyses are now mentioned in Section 2.3 and the results of hourly vs. daily closure are discussed in Section 3.2 with Figure 4 being part of the supplementary material.

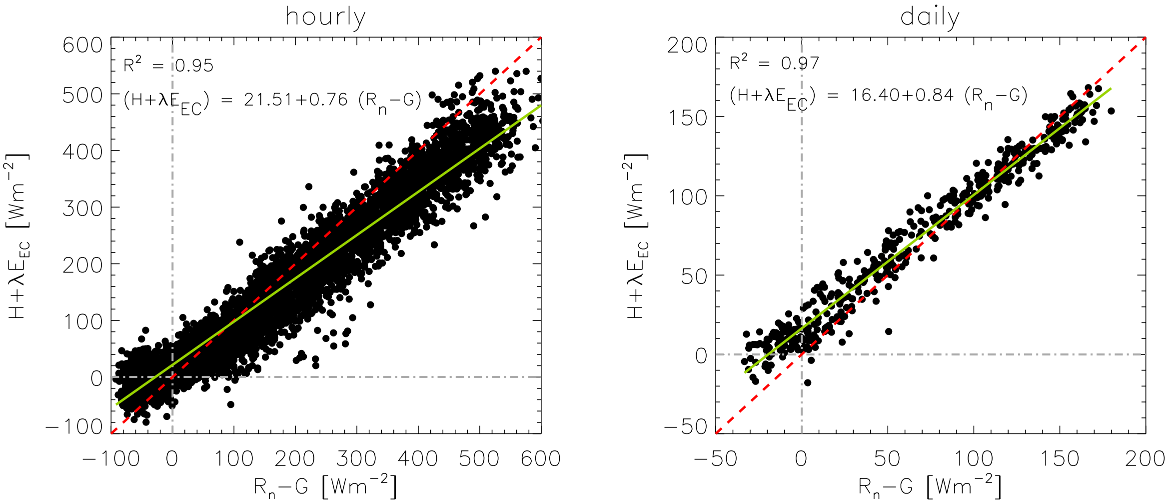

Figure 4: Energy balance closure as evaluated from the ordinary least squares regression based on (left) hourly data and (right) daily data. Only days where maximal five of the hourly values were gapfilled are considered. The dashed red line indicates the 1:1 line, the green line the regression line.

In order to take this into account in the lysimeter comparison, we provide a modified Figure 3 of the manuscript in the supplementary material, showing E_{EC_BOWEN} from the daily force-closure along with the hourly force-closure and the lysimeter estimates (see Figure 5 below).

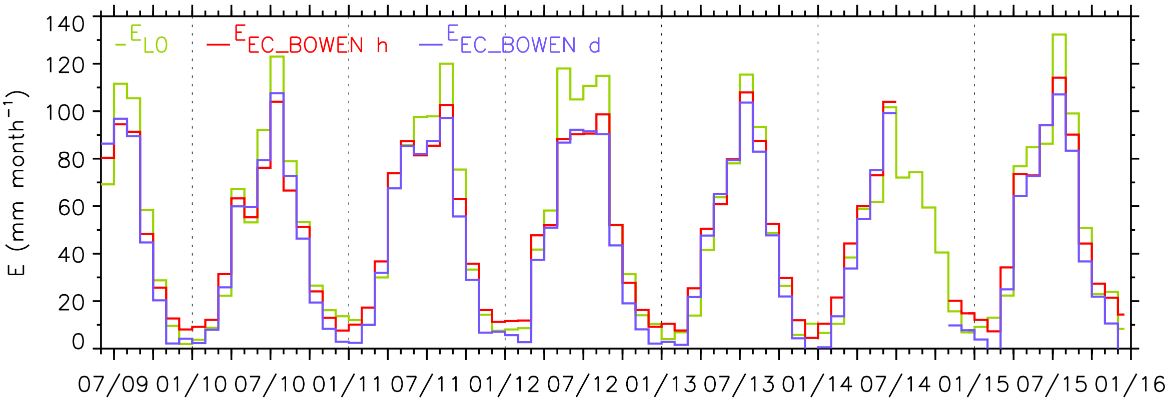

Figure 5: Monthly values of the different evapotranspiration estimates for the time period June 2009 to December 2015. E_{L0} denotes lysimeter evapotranspiration with values set to zero during hours with rain, E_{EC_BOWEN} EC-based evapotranspiration force-closed on hourly ($E_{EC\_BOWEN\ h}$) and daily ($E_{EC\_BOWEN\ d}$) time scale according to the Bowen ratio.

Overall, the daily force-closure leads to a reduction of EC evapotranspiration compared to the one based on the hourly force-closure. The existing underestimation in summer based on hourly force-closure thus becomes slightly larger, while the rather positive EC bias in winter turns into a predominantly negative bias. These results are mentioned in Section 3.3 and Figure 5 is part of the supplementary material.

Specific comments: Lines 42-45: Although you cannot review all of the previous studies, you should at least discuss how your study builds on them (new study region, longer time record, etc.). Some discussion of Alfieri et al.'s differences (and relationship to heterogeneity in vegetation), would also be good.
We expanded the discussion of existing studies in the introduction. Also, we more strongly highlight the fact that our study as compared to previous ones is based on a multi-year

comparison of the two measurement methods. And that it is based on data of a non-irrigated site in a temperate humid climate, while many previous studies were carried out in irrigated agroecosystems in semi-arid or arid climate.

Line 105: "Relatively-large" is subjective here. Your lysimeter still only has a surface area of 0.8 m. I have worked with large weighing lysimeters with 8 m2 surface area.
The Rietholzbach lysimeter has a surface area of 3.14 m^2 (i.e., radius of 1 m) and a total depth of 2.5 m, which according to literature can be considered as a large lysimeter. Nevertheless, we omitted the term "relatively large" and changed the sentence to: "This size of the vessel ensures a higher quality of the measurements"

Line 185: This is organized a bit awkwardly as you present your EC instrumentation and theory, then energy budget theory, then the G and Rn measurements in section 2.5. Might be good to add (see section 2.5) somewhere with G in this sentence so readers don't go back looking for details about the Rn and G instruments.
This noted in the subsequent sentence: "Details on the measurements of R_n and G are given in Section 2.5." In addition, we added a "see below" after mentioning R_n and G.

Additional references:

Xuhui Lee, 1998. On micrometeorological observations of surface-air exchange over tall vegetation. Agricultural and Forest Meteorology, 91, 39-49, doi: 10.1016/S0168-1923(98)00071-9.

Paw U, K. T., Baldocchi, D. D., Meyers, T. P., Wilson, K. B., 2000. Correction of Eddy-Covariance measurements incorporating both advective effects and density fluxes. Boundary-Layer Meteorol., 97, 487-511, doi: 10.1023/A:1002786702909.

Casso-Torralba, P., Vilà-Guerau de Arellano, J., Bosveld, F., Soler, M. R., Vermeulen, A., Werner, C., Moors, E., 2008. Diurnal and vertical variability of the sensible heat and carbon dioxide budgets in the atmospheric surface layer. Journal of Geophysical Research, 113, D12119, doi: 10.1029/2007JD009583.