hess-2016-252: A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration by Hirschi et al.

Reply to Referee #1
We thank Referee #1 for the positive feedback to our manuscript. We’re happy that the value of our long-term measurements is recognized and that the presented inter-comparisons and analyses are well-received.

Reply to Referee #2
We thank Referee #2 for the comments (in italic in the following) and will address them point-by-point below.

This manuscript reports on a long monitoring of evapotranspiration measurements obtained by different techniques. The topic is not novel, even if it benefits of the use of the lysimeter observations.
To our knowledge, this study is the first presenting a multi-year comparison of EC and lysimeter measurements. The existing comparisons are strongly limited by the length of the analyzed time periods (generally not available beyond a few months). Here, we are in the rare situation of having both techniques (lysimeter and EC) operated in parallel at the same location over a multi-year time period. We are not aware of other studies having parallel measurements with these two techniques over such a long time period. This allows us to perform an unprecedented comparison between two well established techniques and shed light on the quality and reliability of monitored data based on these methods.

The paper is generally well written and organized and easily to understand, except for the Figures, which should report on their caption the meaning of each of the used acronym. In the follow, I have included all the modification I retain necessary before the paper publication.
We adjusted the captions of the figures with the explanation of the acronyms.

Title: please avoid the use of the dash in between eddy and covariance.
We modified the title accordingly.

Introduction: the authors should expand the number of studies similar to that conducted by them.
We expanded the number of mentioned studies and their discussion in the introduction. Also, we more strongly highlight the difference to previous studies.

Methods and data
Figure 2 is not easy to interpreter; Figure 2 should include the map of the site, with the footprint area of the micrometeorological EC tower.
We added an aerial online map of the site in order to facilitate the interpretation of Figure 2 (also available here: https://s.geo.admin.ch/6de2dcf3b5). Note that the lowest measurement level (2 m) was chosen as its source area is smallest and closer to the lysimeter (see lines 172-174). However, up to 10% of the measurements are potentially affected by obstacles (trees and a farmhouse) in the area, whereas only 1% of the measurements within the main wind direction (i.e., from west, see Figure 2 of the manuscript) are potentially influenced (Peter, 2011, based on the footprint model of Kljun et al., 2004). This has been noted in Section 2.3.
Eddy Covariance measurements: it is not clear if the sonic and the IRGA have the same point of measurements; please explain.

As mentioned on lines 190/191, the IRGA and the sonic have a horizontal separation of 0.2 m. Thus, they have not the same point of measurement, as the IRGA would extensively disturb the sonic signal and make measurements impossible.

Is the level of measurement consistent with the presence of the roughness sub-layer; how the authors have determined it?
A roughness sub-layer thickness of up to three times the canopy height, i.e. maximum of 1.2 m in the present study, is assumed according to the well-established rule of thumb mentioned in the provided literature. We noted this on lines 177/178:
“Note that this level is well above the vegetation height (mostly below 15 cm, maximum 40 cm), and clear of the roughness sublayer (estimated three times the canopy height; see Kaimal and Finnigan, 1994; Foken, 2008).”

At the end, which method (among those indicated) was used to correct the energy balance closure? Details should be added. In this case, the authors refer too much to studies conducted by others.
All three energy balance closure approaches are applied (see line 232) with the aim to provide the possible range of $E_{EC}$. Accordingly, we show the data from all three methods in the corresponding figures and tables (denoted $E_{EC,BOWEN}$, $E_{EC,H}$ and $E_{EC,E}$). The aim of the paper is not to evaluate the energy balance closure problem in detail, but to inter-compare the two measurement techniques (lysimeter and EC) based on the rare fact of having both techniques operated at the same location over a multi-year time period.

Catchment water balance measurements: The reason why the authors have applied this simplified hydrological water balance should be better justified. The temporal resolution of these estimates does not match with the EC resolution. The used approach appears too much simplified and the reason why some variables have been neglected should be indicated.
As mentioned in Section 2.4 (line 247), the catchment water balance is able to provide an additional estimate of the evapotranspiration amount over a hydrological year. It is used since it is a valuable independent way to evaluate the local-scale lysimeter measurements over the long measurement period. As also mentioned (lines 250-253), it neglects storage changes from year to year. This assumption generally only holds for long-term averages ($\geq 1$ year), thus it is unfortunately not possible to apply the catchment water balance for higher temporal resolution. Of course, the higher temporal resolution data from the lysimeter and EC measurements have been accumulated over the hydrological year for these comparisons.

How $Q_{C}$ was obtained?
$Q_{C}$ is the runoff captured at the catchment outlet at the gauge “Rieholz-Mosnang” (see lines 261/262). In order to clarify this, we adjusted the corresponding sentence to:
“Catchment runoff ($Q_{C}$) is captured at the catchment outlet …”

Additional measurements at the site: How many soil temperature probes were used?
Sometimes the lack of EB closure may depend on the correct estimation of the G term. Please include specifics.
Three soil temperature sensors were used, which are installed at the same locations as the heat-flux plates, and which were also averaged for the estimation of the surface soil heat flux. This has been clarified on lines 276/277. To include more specifics of $G$, we now provide an analysis of the heterogeneity of the surface and 5-cm soil heat fluxes and of soil temperature
in the supplementary material (based on the three heat-flux plates and three soil temperature sensors, see reply to Reviewer #3 below).

**Results**
The results are mainly affected by the choice of the authors to compare data with different temporal resolution of acquisition. This may be only justified from their wish to provide a suggestion to the readers. But this motivation does not appear into this paper.

We do not understand what the reviewer means with this comment, which is not justified given the analyses performed. We provide comparisons on annual down to hourly time scales in order to document the data from the two measurement techniques on various time scales. Contrary to what is suggested by the reviewer, the temporal resolutions of the compared time series are consistent. Hence the underlying differences are not related to the temporal resolution of the data acquisition but to the differences in measurement techniques.

*The BR correction procedure should be explained.*
The Bowen ratio correction is a commonly used approach to correct EC data (see e.g., Twine et al. 2000) and we refer to corresponding papers on line 239. We expanded the respective description as follows:

“Approach (i) ($E_{EC, BOWEN}$) is a commonly used assumption in the literature (e.g., Twine et al., 2000, Jaeger et al., 2009; Jung et al., 2010). It assumes that the Bowen ratio is correctly measured by the EC method so that $\lambda E$ and $H$ can be adjusted to balance equation (3).”

*In such a work, I would have expected to find at least the analysis of unstable and stable conditions of the EC fluxes.*

This analysis would be interesting for characterizing the planetary boundary layer and comparing the micrometeorological conditions to other sites. While we acknowledge that the analysis of stability is performed for corresponding studies focusing on EC measurements, we want to point out that this paper focuses on the comparison of different evapotranspiration estimates, which are defined here as upward fluxes (see lines 149 and 398). Thus, in the lysimeter vs. EC comparisons evapotranspiration from EC measurements is based on (neutral to) unstable conditions only (we noted this in the first paragraph of Section 3.3). In our opinion an analysis of the atmospheric stability would not add to the focus of the article.

*Finally, I have not evinced any message from this work, to improve knowledge of the scientific community. For these reasons, I only suggest Major Revision to this paper, to be reconsidered for publication.*

As mentioned, the rare fact of having lysimeter and EC measurements operated over a multi-year time period is one of the values of this paper. Existing comparisons are strongly limited by the length of the analyzed time periods (generally not available beyond a few months). Moreover, they were often carried out in irrigated agroecosystems of semi-arid or arid climate. Here we present data from a non-irrigated site in temperate humid climate. Presenting this data to the scientific community is indeed worthwhile and the presented inter-comparisons shed light on the quality and reliability of monitored data from the different methods at different temporal scales and over an extended time period.

Additional references:

Reply to Referee #3

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 reply to Reviewer #2 above).

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 \( F_z = \overline{w_z}(\overline{m_z} - \langle m \rangle) \) 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. \( F_z \) 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.
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 ($E_{LO}$) 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.

<table>
<thead>
<tr>
<th>Wind Speed</th>
<th>High Wind Speed</th>
<th>Low Wind Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSD</td>
</tr>
<tr>
<td>East</td>
<td>0.79</td>
<td>0.090</td>
</tr>
<tr>
<td>West</td>
<td>0.77</td>
<td>0.081</td>
</tr>
<tr>
<td>South</td>
<td>0.76</td>
<td>0.017</td>
</tr>
<tr>
<td>All</td>
<td>0.78</td>
<td>0.084</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th>Vapor Pressure Deficit</th>
<th>High Vapor Pressure Deficit</th>
<th>Low Vapor Pressure Deficit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSD</td>
</tr>
<tr>
<td>All</td>
<td>0.76</td>
<td>0.088</td>
</tr>
<tr>
<td>Day</td>
<td>0.67</td>
<td>0.107</td>
</tr>
<tr>
<td>Night</td>
<td>0.06</td>
<td>0.031</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Wind</th>
<th>All</th>
<th>East</th>
<th>West</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>Rel. Bias</td>
<td>Number</td>
<td>$R^2$</td>
</tr>
<tr>
<td>All</td>
<td>0.80</td>
<td>-0.029</td>
<td>30002</td>
<td>0.81</td>
</tr>
<tr>
<td>Day</td>
<td>0.75</td>
<td>-0.001</td>
<td>16154</td>
<td>0.77</td>
</tr>
<tr>
<td>Night</td>
<td>0.04</td>
<td>-0.374</td>
<td>13847</td>
<td>0.03</td>
</tr>
</tbody>
</table>
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).](image)

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. $E_{LO}$ 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 m$^2$ 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 $Rn$ and $G$ are given in Section 2.5.”

Additional references:


Additional changes

Figure 7 of the manuscript has been changed in order to only include measured values both from EC and from the lysimeter (before this was only the case for the EC data).
The energy balance closure analyses are now based on measured values only (excluding gap-filled data). Thus, the respective numbers in Section 3.2 and the corresponding Figure 4a have been updated. For consistency, also Figure 8 is now only based on measured values. Note that this change has only minor effects.

We now consistently use “precipitation” instead of “rain” throughout the manuscript.
A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration

Martin Hirschi¹, Dominik Michel¹, Irene Lehner¹,², and Sonia I. Seneviratne¹

¹ Institute for Atmospheric and Climate Science, ETH Zurich, Universitätstrasse 16, 8092 Zürich, Switzerland.
² Present address: Centre for Environmental and Climate Research (CEC), University of Lund, Lund, Sweden.

Correspondence to: Martin Hirschi or Sonia I. Seneviratne (martin.hirschi@env.ethz.ch, sonia.seneviratne@env.ethz.ch)
Abstract. Accurate measurements of evapotranspiration are required for many meteorological, climatological, ecological, and hydrological research applications and developments. Here we examine and compare two well-established methods to determine evapotranspiration at the site level: lysimeter-based measurements ($E_L$) and eddy-covariance (EC) flux measurements ($E_{EC}$). The analyses are based on parallel measurements carried out with these two methods at the research catchment Rietholzbach in northeastern Switzerland, and cover the time period June 2009 to December 2015. The measurements are compared on various time scales, and with respect to a 40-year lysimeter-based evapotranspiration time series. Overall, the lysimeter and EC measurements agree well, especially on the annual time scale. On that time scale, the long-term lysimeter measurements also correspond well with catchment water-balance estimates of evapotranspiration. This highlights the representativeness of the site-level lysimeter and EC measurements for the entire catchment despite their comparatively small source areas and the heterogeneous land use and topography within the catchment. Furthermore, we identify that lack of reliable EC measurements using open-path gas analyzers during and following precipitation events (due to limitations of the measurement technique under these conditions) significantly contributes to an underestimation of $E_{EC}$ and to the overall energy balance gap at the site.

Keywords. Evapotranspiration, lysimeter, eddy covariance, catchment water balance, measurements
Evaporation $E$ from land, also termed evapotranspiration, is an essential contributor to the water and energy balances on continents. It returns about 60% of the precipitated water on land back to the atmosphere and also uses up more than 50% of all net radiation available on land (e.g., Oki and Kanae, 2006; Trenberth et al., 2009; Jung et al., 2010; Seneviratne et al., 2010; Wang and Dickinson, 2012). In addition, it is coupled to the carbon dioxide (CO$_2$) uptake by vegetation (e.g., Farquhar and Sharkey, 1982; Sellers et al., 1996; Claïs et al., 2005; Reichstein et al., 2013), which implies important links between carbon and water cycles. Furthermore, evapotranspiration is related to other nutrient cycles such as the nitrogen cycle (Larcher, 2003).

Approaches to measure or estimate evapotranspiration are diverse and can include ground observations, remote sensing-based estimates, diagnostic techniques, as well as modelling and reanalyses (e.g., Seneviratne et al., 2010; Jiménez et al., 2011; Müller et al., 2011; Wang and Dickinson, 2012). Despite their relative scarcity, the best-established reference measurement remains ground observations, which can be for example either performed with the lysimeter technique commonly used in hydrology (e.g., Maidment, 1992; Rana and Katerji, 2000; Seneviratne et al., 2012), or the eddy-covariance flux measurement technique (EC) established in micrometeorology (e.g., Baldocchi et al., 2001; Aubinet et al., 2012). Both techniques hold specific intrinsic limitations (see Section 2).

Unfortunately, long-term parallel measurements of evapotranspiration with different techniques are rare. Many studies are limited in terms of number of methods and/or length of analyzed time period (e.g., Schume et al., 2005; Kosugi and Katsuyama, 2007; Castellví and Snyder, 2010; Wang and Dickinson, 2012 and references therein). Reported differences between lysimeter and EC measurements from these short-term comparisons amount from a few percent up to 30%, with EC evapotranspiration being mostly lower than the lysimeter-based values (e.g., Chávez et al., 2009; Ding et al., 2010; Gebler et al., 2015). It is also worth mentioning the BEAREX08 field experiment, within which several methods of determining evapotranspiration were evaluated at different spatial scales for one vegetation period (Evett et al., 2012a,b) and substantial differences between EC and lysimetry were reported (Alfieri et al., 2012). In particular, the impact of surface heterogeneity (i.e., vegetation density) on the uncertainties of both lysimeter and EC measurements, as well as the influence of advective fluxes and energy balance closure deficits on the discrepancy between the two methods were investigated. Over 35% of the discrepancy could be attributed to differences in vegetation, while for the rest imperfect energy balance closure, advective effects and sensor-related measurement uncertainties were responsible.

The purpose of the present study is to compare lysimeter- and EC-based measurements of evapotranspiration in the pre-alpine Rietholzbach catchment, which is characterized by a unique hydroclimatological record, including lysimeter...
measurements since 1976 (Seneviratne et al., 2012). As compared to numerous previous studies (see above), which were carried out in irrigated agroecosystems in semi-arid or arid climate, the presented inter-comparison is based on data of a non-irrigated environment in a temperate humid climate. In 2009, EC sensors were installed, thus allowing an extensive, multi-year comparison between the two techniques (as compared to the shorter-term comparisons of previous studies).

Furthermore, we use the catchment-wide water balance as an additional constraint for estimated evapotranspiration on the yearly time scale. Hence, we compare three approaches to measure or estimate evapotranspiration, which vary both in temporal resolution (from minutes to a year) and spatial scale (from m$^2$ to km$^2$). This allows us to evaluate (i) the correspondences between the two well-established local-scale evapotranspiration measurement techniques, (ii) the representativeness of these local-scale measurements for the catchment, and (iii) the quality of the EC measurements under the considered conditions at the site.

This article is structured as follows: The methods and data employed in this study are presented in Section 2. Section 3 shows the resulting evapotranspiration estimates by the different techniques and on different time scales. Section 4 discusses the results, and the summary and conclusions of this study are provided in Section 5.

2 Methods and data

2.1 Site description and catchment characteristics

The measurements considered in this study were conducted at the hydrometeorological research catchment Rietholzbach in northeastern Switzerland (47.37°N, 8.99°E, 795 m asl; see Seneviratne et al., 2012 for an overview of the site). The hilly, pre-alpine catchment (elevation range: 682-950 m asl) drains an area of 3.31 km$^2$ and is a headwater catchment of the Thur river. The region is characterized by a temperate humid climate with a mean air temperature $T_{air}$ of 7.1 °C and ample precipitation $P$ with a mean annual sum of 1438 mm yr$^{-1}$ (data basis 1976-2015, Figure 1a,b). Net radiation $R_n$ exhibits a clear seasonal pattern with on average 105 W m$^{-2}$ in summer and -7 W m$^{-2}$ in winter (data basis 2000-2015, Figure 1c).

Predominantly weak winds (77% below 2 m s$^{-1}$) blow along the east-west orientation of the valley (Figure 2). Catchment runoff $Q_C$ is strongly related to subsurface storage (Teuling et al., 2010a) and shows an annual mean of 104 L s$^{-1}$, which corresponds to 991 mm yr$^{-1}$. It displays lowest values in summer and its peak value during snowmelt in March (data basis 1976-2015, Figure 1d). The conglomerate Nagelfluh, the main parent rock type, originates from the Würm glaciation. The soil type and depth exhibit a high spatial variability. Overall, shallow regosols dominate on steep slopes, deeper cambisols are found in flatter areas, and gley soils are located in the vicinity of small creeks. Land use underwent no major changes since the start of the measurements in late 1975 and is highly related to the topography. On slopes and along creeks, in about one fourth of the area, forest dominates. Otherwise the area is used as grassland and partially as pasture. The catchment is only sparsely populated.
Most measurements considered in this article are conducted at the site “Büel”, which is located in a grassland area next to the valley bottom in the upper part of the Rietholzbach catchment (see Figure 2 and https://s.geo.admin.ch/6de2dcf3b5). The on-going measurements include standard meteorological and hydrological variables such as air temperature, precipitation, air humidity, radiation, soil moisture, runoff, and ground water level. Evapotranspiration measurements are provided by a lysimeter and an eddy flux tower (see Figure 2 for an overview on the setup of these measurements). Further details on the relevant instrumentation for this study are given in the following sections. Seneviratne et al. (2012) provide an overview of the characteristics of the catchment and of measurements at the site. For more general information about the catchment we also refer to www.iac.ethz.ch/url/rietholzbach.

2.2 Lysimeter measurements

Lysimetry is a well-established technique to measure evapotranspiration (e.g., Maidment, 1992; Rana and Katerji, 2000; Goss and Ehlers, 2009; Meissner et al., 2010; see also Seneviratne et al., 2012 for a recent overview). Lysimeters are vessels containing a soil column in near-natural condition. Weighing lysimeters allow the quantitative measurement of water changes within the soil column and thus, in combination with precipitation and lysimeter seepage measurements, also the estimation of evapotranspiration.

Formally, the lysimeter evapotranspiration $E_L$ (mm) within a given time interval $\Delta t$ (here one hour) is estimated from the initial weight $W_t$ minus the final weight $W_t^{+1}$ (both in kg), the precipitation $P$ (mm), and lysimeter seepage $Q_L$ (mm) at the vessel bottom:

$$E_L = \frac{W_t - W_t^{+1}}{\rho_w \pi r^2} + P - Q_L,$$  \hspace{1cm} (1)

where $\rho_w$ stands for the density of water (kg m$^{-3}$) and $r$ (m) for the radius of the lysimeter. For the comparison with eddy covariance measurements, we derive a parallel time series “$E_{L0}$” in which $E_L$ is set to zero during hours with precipitation $P \geq 0.1$ mm, 15.4% of data), as no reliable data is available from the EC measurements in those cases (see Section 2.3).

The Rietholzbach lysimeter has a surface area of 3.14 m$^2$ (radius of 1 m) and a total depth of 2.5 m including a gravel filter layer at the bottom. This size of the vessel ensures a higher quality of the measurements (see Seneviratne et al., 2012 and references therein). The lysimeter weight is measured with three load cells and a resolution of 100 g, which corresponds to a water column of approximately 0.03 mm. The surface is covered by grass of similar species composition and treated...
according to the surrounding grassland (same cutting scheme, but synthetic fertilization instead of slurry; see also Figure 2). At its installation at the site “Büel” in late 1975 the lysimeter was backfilled with a typical gleyic cambisol. Seepage at the lysimeter lower boundary is measured by a tipping bucket with a volume of 50 ml, i.e., with a resolution of 0.02 mm (Gurtz et al., 2003). Following the recommendations of the World Meteorological Organization (WMO, 2008), precipitation data were not derived from the lysimeter measurements but were taken from a standard tipping bucket (see Section 2.5).

A key requirement for the accurate estimation of local evapotranspiration is the representativeness of the lysimeter for the surrounding area in terms of soil conditions, vegetation composition, and treatment. Major drawbacks are the existence of the vessel and its specific design (Allen et al., 2011; WMO, 2008). At the investigated site, this implies the main following limitations:

(i) The lateral water transport is not contributing to the lysimeter water storage dynamics. This point is assessed as being relatively negligible, as the lysimeter is located in a flat area close to the valley bottom and surrounded by a slight and uniform slope. Thus, potential lateral in- and outflow to the investigated soil volume are assumed to be equal.

(ii) There is no connection to the groundwater. This may become potentially important under drought conditions (e.g., Rana and Katerji, 2000; Seneviratne et al., 2012) even for a grass-covered lysimeter in a temperate humid climate.

(iii) Drainage occurs by gravitation only as soil suction is not artificially reproduced within the vessel.

(iv) Time periods with snow cover have to be analyzed with special care as snow drift induced by wind as well as snow bridges to the surrounding can falsify the weight measurement.

Despite these issues, Seneviratne et al. (2012) show that the Rietholzbach lysimeter seepage and catchment runoff display very similar monthly dynamics, which suggests to a first approximation, that the lysimeter is well representative for the entire catchment despite the scale discrepancy and mentioned limitations. The largest discrepancies between lysimeter seepage and catchment runoff are found in March, most likely linked to a higher spatial variability of hydrological processes in that month, due to snow melt and the onset of the growing season.

Lysimeter data analysed in this study cover the time period 1 June 2009 until 31 December 2015 (start being restricted by the availability of the EC measurements, see Section 2.3). In addition, we refer to the climatological lysimeter time series dating back to 1976 (see Section 2.5). Evapotranspiration is calculated in hourly time steps according to equation (1) and taking into account the weight changes due to management activities. Missing values in lysimeter weight change $W_t - W_{t+1}$ (<0.1% of data) are filled by a linear interpolation when the gaps were short and no precipitation occurred. For lysimeter seepage $Q_L$ missing values (<0.1% of data) are filled manually preserving the actual seepage pattern. Evapotranspiration is defined here as an upward flux, i.e., comprising positive values only, as the lysimeter accuracy does not allow to resolve dew formation.
Yet, equation (1) can result in negative $E_L$ values, because the measurements entering the calculation are based on instruments with differing resolutions, and because they can be biased due to sensor uncertainty. The latter is in particular the case for the precipitation measurements, which are often biased due to an undercatch (e.g., Sevruk, 1982; Adam and Lettenmaier, 2003). Days comprising negative $E_L$ are thus treated as described in Jaun (2003) to eliminate such negative values, in consistence with the scheme used for the climatological data series dating back to 1976 (see Section 2.5). It takes into account the amount and predominant sign of $E_L$ during such days. The method mainly affects the winter period, when it leads to a reduction of the overall amount of positive $E_L$ as compared to e.g., simply setting all negative $E_L$ values to zero (not shown), as the occurrence of negative $E_L$ is increased during this season. Based on measured values a threshold for maximum realistic $E_L$ of 0.2 mm h$^{-1}$ during nighttime (global radiation $R_{al} < 10$ W m$^{-2}$) and periods with snow cover (albedo $\alpha > 0.5$) is applied. In addition, a limitation of $E_L$ is defined as a function of $R_{al}$. The subsequent gap filling is conducted in two steps: (i) missing nighttime values (0.7% of data) are set to zero, and (ii) for missing daytime values (0.8% of data) a linear regression with global radiation $R_{al}$ ($R^2$: 0.67) is applied.

### 2.3 Eddy-covariance measurements

The eddy-covariance method estimates the vertical mass flux of water vapour ($E_{EC}$) exchanged by an ecosystem based on fast measurements of vertical wind velocity $w$ (m s$^{-1}$) and specific humidity $q$ (kg m$^{-3}$), respectively on their turbulent fluctuations (denoted by a prime):\[ E_{EC} = \overline{w'q'}. \] (2)

$E_{EC}$ is measured at the Rietholzbach catchment since late May 2009. The measurements are conducted on a 9 m flux tower, installed at the site “Büel” (see Figure 2 and https://s.geo.admin.ch/6de2dcf3b5) and equipped on three levels with an ultrasonic anemometer thermometer (sensor type CSAT3, Campbell Scientific Inc., USA; hereafter referred to as “sonic”).

On the bottom and top levels, an open-path CO$_2$/H$_2$O infrared gas analyzer (sensor type Li-7500, LI-COR Biosciences, USA; hereafter referred to as “IRGA”) complete the set-up. The instruments are operated at 10 Hz and data is saved with a CR3000 data logger (Campbell Scientific Inc., USA). The present study is based on data obtained from the sensors at the lowest level (2 m above ground), as their source area is smallest and closer to the lysimeter (see Figure 2), and therefore, they experience more homogeneous physical environmental conditions. However, up to 10% of the measurements are potentially affected by obstacles (trees and a farmhouse) in the area, whereas only 1% of the measurements within the main wind direction (i.e., from west, see Figure 2) are potentially influenced (Peter, 2011, based on the footprint model of Kljun et al., 2004). Note that this level is well above the vegetation height (mostly below 15 cm, maximum 40 cm), and clear of the roughness sublayer (estimated three times the canopy height; see Kaimal and Finnigan, 1994; Foken, 2008). In this study we consider data from the time period 1 June 2009 until 31 December 2015.
To enable the comparison with the lysimeter estimates, the statistics were calculated on an hourly time step following the methodology described in e.g., Lee et al. (2004) or Aubinet et al. (2012). This includes a time lag correction of \( w \) and \( q \) by maximization of their covariance, the application of the planar fit method after Wilczak et al. (2001) for the coordinate rotation, spectral correction (Moore, 1986), conversion of the buoyancy flux into the sensible heat flux (Schotanus et al., 1983), and correction of density fluctuations (Webb et al., 1980). As open-path IRGAs are not reliably measuring when water is accruing on the optical elements, \( \Delta E_{EC} \) is explicitly set to zero during hours with precipitation \( (P \geq 0.1 \text{ mm}, 15.4\% \) of the data).

Figure 2 displays the location of the tower with respect to the lysimeter and the radiation measurements, together with the frequency of wind direction and velocity at the location. The horizontal separation distance between sonic volume and IRGA volume amounts to 0.2 m both laterally and longitudinally, with the IRGA being situated west of the sonic (not shown). EC data is masked when the tower and the IRGA are in the upwind direction of the sonic volume (i.e., from 310 to 50 degree, red sector) in order to avoid impacts on the measured turbulent fluxes. This is the case for 10.6\% of the data.

Evapotranspiration \( E_{EC} \) is related to the surface energy balance as follows:

\[
R_n = G = H + \lambda E_{EC}, \tag{3}
\]

where \( R_n \) refers to the net radiation, \( G \) is the surface soil heat flux (see below), \( H \) stands for the sensible heat flux, and \( \lambda E_{EC} \) stands for the latent heat flux, where \( \lambda \) (J kg\(^{-1}\)) is the latent heat of vaporization. Details on the measurements of \( R_n \) and \( G \) are given in Section 2.5. The storage of energy between the surface and the measurement height is neglected in the analyzed measurements as these are performed at 2 m above short grassland (vegetation height mostly below 15 cm, maximum 40 cm). However, it is not negligible for tall vegetation (e.g., Foken et al., 2006). In addition, effects of diurnal storage changes can be averaged out when considering daily instead of hourly energy balances (e.g., Leuning et al., 2012; Anderson and Wang, 2014).

Advective fluxes are also not considered in equation (3). This effect can contribute to a non-closure of the surface energy balance (e.g., Leuning et al., 2012, see also below). The vertical component of advection of latent and sensible heat (see e.g., Paw U et al., 2000; Casso-Torralba et al., 2008) can be assessed at the site following the notation of Lee (1998), which is based on the average vertical gradient of moisture or temperature multiplied by the mean vertical wind speed at a specific level. For a quantitative estimate of the horizontal component of advection, measurements are not available at the site and its surroundings. Possible reasons for horizontal advection include slope drainage in complex terrain and heterogeneous land cover (e.g., Katul et al., 2006). Concerning the first reason (slope drainage), the wind from the south-facing valley slope is masked in the whole analyses as it includes the tower (see above). Concerning the impact of surface heterogeneity, we estimate the potential effect on the energy balance closure by separating the analyses into three wind sectors (i.e., east, west
and south wind directions). While the west sector (i.e., the main wind direction) features homogeneous land cover and horizontal advection should thus not be relevant, the east sector is potentially impacted by a small street and a farmhouse (see Figure 2).

For the latent heat flux $\lambda E_{EC}$ the same data constraints are applied as for lysimeter evapotranspiration $E_L$, i.e., during nighttime conditions, periods with snow cover, and limitation by $R_{sd}$ (see Section 2.2). Under the present generally humid climate conditions at Rietholzbach, net radiation $R_n$ is the main driver and limiting factor for $\lambda E_{EC}$ (Teuling et al., 2010b; Seneviratne et al., 2012). Thus, gaps in the $\lambda E_{EC}$ time series (31.1% of data) are filled by a linear regression ($R^2$: 0.90) of these two variables. However, it should be noted that the simple regression with radiation could lead to errors when evapotranspiration is constrained by soil moisture (e.g. Seneviratne et al., 2010). Overall, the relation between the $\lambda E_{EC}$ and the lysimeter time series is not changed by the gap filling (not shown).

As commonly observed with using EC data (e.g., Twine et al., 2000; Wilson et al., 2002; Franssen et al., 2010), the energy balance is not closed at the investigated site (see Section 3.2), i.e., the available energy ($R_n - G$) is generally higher than the sum of the turbulent fluxes ($H + \lambda E_{EC}$). This known issue of the EC method is extensively discussed in the literature (e.g., Mahrt, 1998; Foken, 2008; Aubinet et al., 2012; Leuning et al., 2012). It is important to address this issue also in light of the use of EC data for model validation (e.g., Jaeger et al., 2009).

Several approaches can be used to force-close the energy balance. Here we apply three different simple approaches to enforce the energy balance closure on an hourly basis, assigning the gap to:

1. both $H$ and $E_{EC}$ according to the Bowen ratio $\beta$ ($E_{EC, Bowen}$)
2. sensible heat flux only ($E_{EC,H}$)
3. latent heat flux only ($E_{EC,E}$).

Due to weak turbulent conditions, small turbulent fluxes, and a poor definition of the Bowen ratio during nighttime, the approaches are only applied to daytime values ($R_{sd} \geq 10$ W $m^{-2}$). Approach (i) ($E_{EC, Bowen}$) is a commonly used assumption in the literature (e.g., Twine et al., 2000; Jaeger et al., 2009; Jung et al., 2010). It assumes that the Bowen ratio is correctly measured by the EC method so that $\lambda E_{EC}$ and $H$ can be adjusted to balance equation (3). Approaches (ii) and (iii) represent two extreme assumptions but they are useful as they indicate the entire range of possible energy balance options (given that $R_n$ and $G$ are correctly estimated and no other fluxes (e.g., advection) or storage terms are of importance, see also Section 4.2). For comparison, approach (i) is also applied on daily time scale (i.e., based on daily aggregated fluxes).
2.4 Catchment water balance measurements

The catchment water balance integrates its components over the entire catchment area over longer time periods. While we focus here on the comparison of the lysimeter and EC evapotranspiration measurements, such catchment water-balance estimates provide an additional reference to evaluate the local-scale techniques. Using this approach, the evapotranspiration $E_C$ is estimated as the difference between precipitation $P$ and catchment runoff $Q_C$ (all in mm):

$$E_C = P - Q_C.$$

(4)

This approach implies that the change in catchment storage over the given time interval is zero. This assumption generally only holds for long-term averages ($\geq 1$ year). Although year-to-year variations in storage cannot be fully excluded (e.g., Seneviratne et al., 2012), it can be assumed to yield a reasonable estimate for hydrological years (October to September in Switzerland). In addition, catchment precipitation is estimated here using one precipitation gauge only (thus assumed to be spatially representative). This approach also assumes that all water is leaving the catchment through the discharge gauge at the catchment outlet (see below). Previous analyses suggest that both conditions are reasonably met for the study catchment (Gurtz et al., 2003; Seneviratne et al., 2012).

Precipitation data is taken from the standard tipping bucket (for details see Section 2.5). As the catchment evapotranspiration $E_C$ is known to suffer from unrealistic negative values during winter (Lehner et al., 2010), related to high precipitation undercatch during snow fall (up to 60%, see Gurtz et al., 2003), the precipitation data entering equation (4) is corrected for undercatch (based on Gurtz et al. 2003, Table 1 therein). Catchment runoff ($Q_C$) is captured at the catchment outlet at the gauge “Rietholz-Mosnang”. This gauge is operated by the FOEN (Federal Office for the Environment, Hydrology Division, Berne, Switzerland). More information on the gauge is available on www.hydrodaten.admin.ch/en/2414.html or in FOEN (1996).

2.5 Additional measurements at the site “Büel”

Precipitation is measured by a standard tipping bucket installed at 1.5 m. Data from parallel measurements with a standard tipping bucket at 0 m and a weighing pluviometer at 1.5 m at the same measurement site are used for quality assessment and gap filling. For remaining gaps a regression with data from nearby meteorological stations operated by MeteoSwiss is applied.

Net radiation is derived from separate measurements of all four components of the radiation balance (CM21 and CG4, Kipp&Zonen, NL, all ventilated) at a height of 2 m (see Figure 2).
Soil heat flux is captured with three heat-flux plates (HFP01, Hukseflux, NL) installed at 0.05 m below ground and situated within a 1-m periphery, which are averaged for the analysis. Surface soil heat flux $G$ is determined following Fuchs and Tanner (1968) by calculating the change in heat storage above the sensors. This estimation is performed using the average of three soil temperature sensors (107T, Campbell Scientific Inc., USA) installed at the same locations as the heat-flux plates, as well as soil moisture (TRIME-IT, IMKO GmbH, D) and soil density measurements (Mittelbach et al., 2012). The ensemble of the three different locations for soil heat flux and soil temperature is used to account for the spatial heterogeneity of the soil matrix and thus to obtain surface soil heat flux data, which are spatially representative for the footprint domain of all other measured components of the energy balance.

Note that soil heat flux was not measured during a four-month period from July to October 2014 due to a logger failure. This leads to some gaps in the following comparisons due to the fact that the energy gap corrections (i.e., $E_{EC_{BOWEN}}$ and $E_{EC_{E}}$, see Section 2.3) cannot be applied for this time period. Apart from this period, the soil heat flux and net radiation time series only hold few and short gaps (<0.1% of data), which are filled by a linear interpolation.

All measurements mentioned in Section 2 are on-going. The descriptions refer to the instrumentation for the data since (at least) June 2009. The climatological data series (since 1976) are generally based on varying sensors (types), but have been homogenized over time. More details on the climatological record and respective instrumentation since 1976 can be found in Seneviratne et al. (2012) and in a German-language report (Gurtz et al., 2006).

### 3 Results

#### 3.1 Climate conditions in 2009-2015 compared with long-term climatology

We first assess how the measurements in the study period compare with the long-term climatology to evaluate if the study period from 1 June 2009 until 31 December 2015 is representative for the mean climatological conditions at the site. Figure 1 displays the average monthly meteorological conditions during the study period compared to the long-term climatology with respect to air temperature $T_{air}$, precipitation $P$, net radiation $R_n$, catchment runoff $Q_C$, lysimeter evapotranspiration $E_L$, and lysimeter seepage $Q_L$. The long-term climatological values are derived over the time period 1976-2015, with the exception of net radiation, which is only measured since 2000 at the site.

Temperature (Figure 1a) during the study period ranges on average in the upper part of the distribution based on the climatological data series. This is consistent with the recorded long-term increasing temperature trend in Rietholzbach (Seneviratne et al., 2012) and in Switzerland (OcCC, 2008). The variability within the seven years is similar to the climatology. Precipitation of the study period (Figure 1b) shows high variability with extreme values in November 2011...
(0 mm) and December 2011 (275 mm). Overall, the precipitation data shows that the spring season is often drier and the summer season often wetter in the seven considered years compared to the long-term climatology. Absolute values and variability for \( R_e \) (Figure 1c) are close to the long-term average. Catchment runoff and lysimeter seepage (Figure 1d,f) show high variability within the seven years and compared to the climatological values. \( Q_C \) and \( Q_L \) show a similar behaviour (see also Seneviratne et al. 2012). Lysimeter evapotranspiration \( E_L \) often shows higher summer values (i.e., mostly for June and August) in 2009-2015 compared to the climatology.

Mean precipitation \( P \) of a hydrological year sums up to 1598 mm yr\(^{-1}\) (respectively 1446 mm yr\(^{-1}\) when undercatch is not corrected), whereof about 37% are evaporated as \( E_C \) and about 63% leave the catchment as runoff \( Q_C \) (Table 1). All catchment water balance components show a high year-to-year variability, which is highest for \( Q_C \) with respect to the mean. However, none of the components displays a significant trend over the entire time period (see also Seneviratne et al., 2012 for trends in calendar-year values over the time period 1976-2007). Figure 3 displays the catchment water balance for the hydrological years since 1976/77. The comparison with the lysimeter-based evapotranspiration \( E_L \) suggests that in the long-term mean, the \( E/P \) ratio agrees well with the catchment water-balance approach. The discrepancies between \( E_C \) and \( E_L \) on a year-to-year basis are likely due to non-negligible year-to-year variations in terrestrial water storage (soil moisture, groundwater, snow). In fact, the non-equality of \( (E_L + Q_L) \) vs. \( P \) indicates annual storage variations at the lysimeter, while for \( E_C \) the change in catchment storage over the given time interval is assumed to be zero (see Section 2.4).

The hydrological year 2009/10 is one of the wettest hydrological years in terms of precipitation (+14.5% resp. +18.6% compared to the average, undercatch-corrected and uncorrected values), yet the partitioning of \( P \) into \( Q_C \) and evapotranspiration \( (E_C \) and \( E_L \)) is still close to the long-term average (Table 1). The pattern of the hydrological year 2010/11, in contrast, is different. Precipitation is lower than average (-9.8% resp. -6.9%) for that year, but evapotranspiration \( (E_C \) and \( E_L \)) is up to 1.25 times the average, whereas runoff and seepage \( (Q_C \) and \( Q_L \)) display the lowest values of the entire period (-28.8% and -34.2% respectively compared to the average). It should be noted that the spring 2011 was very dry (e.g., Wolf et al., 2013; Wetter et al., 2014; Whan et al., 2015), which can partly explain these features. Both hydrological years 2009/10 and 2010/11 display amongst the highest evapotranspiration values since the beginning of the measurements \( (E_C \) and \( E_L \)). The year 2011/12 appears rather normal on the catchment scale, but the lysimeter again shows high \( E_L \) (+24.8% compared to the average) and low \( Q_L \) (-18.2%). This is followed by a wetter year 2012/13 with higher \( P \) (+10.8% resp. +10.4%), \( Q_C \) (+23.8%) and \( Q_L \) (+16.6%), but lower \( E_C \) and \( E_L \) (-11.1% resp. -3.9%). The years 2013/14 and 2014/15 finally show again lower than average precipitation (up to -10.5%). For the catchment, this resulted in below normal \( Q_C \) for both years (-12.4% and -14.1%) and close to normal \( E_C \) (+4.2% and -2.7%). The lysimeter on the other hand shows for 2013/14 slightly enhanced \( Q_L \) (+6.0%) while \( E_L \) is close to normal (-4.8%). For 2014/15 it experienced a pronounced drying with high \( E_L \) (+14.7%) and low \( Q_L \) (-16.8%), related to the hot and dry summer of that year (e.g., Scherrer et al., 2016). Overall, the study
period is well covering the climatological variability, including both years with rather extreme conditions as well as years close to average conditions (see Table 1).

### 3.2 Energy balance closure

The energy balance closure as evaluated from the ordinary least squares regression between the hourly estimates of the turbulent fluxes \((H + \lambda E_{EC})\) against the available energy \((R_n - G)\) reaches values of 0.77 for the slope and 18.94 W m\(^{-2}\) for the intercept \((R^2: 0.94, \text{ Figure 4a}; \text{ note ideal closure is represented by an intercept of zero and slope of one}). The ratio of the total amount of the turbulent heat fluxes to available energy amounts to 101.9\%, indicating a surplus of turbulent energy. This is due to mostly slight positive nighttime values of the sum of the turbulent fluxes while available energy displays negative values during night (see Figure 4b). Ignoring nighttime values \((R_{n,d} < 10 \text{ W m}^2\)) the closure ratio amounts to 86.4\%, i.e., the sum of the turbulent fluxes \(H + \lambda E_{EC}\) is generally lower than the available energy \(R_n - G\). The regression between the daytime hourly estimates reveals a slope of 0.80 and intercept of 10.63 W m\(^{-2}\) \((R^2: 0.94\). All these values are in the range of values reported in literature (e.g., Wilson et al., 2002). Note that the analyses presented here are based on measured data only (i.e., excluding gap-filled data) and masked for precipitation and for wind directions impacted by the tower (see also Section 4.2).

Hourly energy balance closure is also compared with daily closure for days where maximally five of the hourly values were gapfilled, which leaves 462 days of valid EC observations (Supplementary Figure 1). The energy balance 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. The increase of the energy balance closure from hourly to daily time scale hints at an effect of diurnal storage variations on hourly time scale, which tend to get averaged out on daily time scale (see Section 2.3).

As mentioned in Section 2.3, advective fluxes are not accounted for in the energy balance equation (3) and can contribute to the imbalance between the turbulent fluxes and the measured available energy. The estimated amount of vertical advection reveals 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 Supplementary Figure 2) compared to the respective average turbulent fluxes (less than 1\%). For horizontal advection, the potential effect on the energy balance closure is estimated 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 in that sector). We focus on daytime here in order to rule out biasing due to the differing distribution of nighttime fluxes among the wind sectors, which are typically small yet can have opposite directions. The results of these analyses reveal that the energy balance closure is rather independent of the wind direction (see Supplementary Figure 3). 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 energy balance closure independent of the
wind direction (in light of spatial homogeneity in the west sector and the small street and the farmhouse in the east sector; see Section 2.3) indicates that horizontal advection is not of great importance at the site.

The mean daily patterns of the energy balance components (Figure 4b) show that nighttime $H$ and $\lambda E_{EC}$ often are of similar small magnitude but opposite sign, resulting in slight positive nighttime values of the sum of the turbulent fluxes and an energy balance closure gap equivalent to about the amount $R_n - G$. The zero-crossing of $R_n$ and $H$ occurs at around 07:00 when $\lambda E_{EC}$ starts to increase as well. $G$ is delayed by about 2 hours. All fluxes have their peak value around 13:00. Available energy is larger than the turbulent fluxes throughout the day. The energy balance closure gap displays a pronounced daily cycle. During nighttime the closure gap is almost constant at around 25 W m$^{-2}$ and the largest closure gap is found around noon. The overall daily cycle of the energy closure gap is approximately symmetric around the noon peak, and generally increases with higher fluxes.

Supplementary Figure 4 displays the daily cycles of surface and 5-cm soil heat fluxes, as well as of soil temperature (see Section 2.5). For the latter two, the average based on the three heat-flux plates and the three soil temperature sensors respectively are shown, 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, along with a minimum and maximum estimate based on the individual sensor locations. 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 surface soil heat fluxes amounts to 6.7 W m$^{-2}$ on the average. 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.

### 3.3 Comparison of the different evapotranspiration estimates

In the following we compare the evapotranspiration estimates on different time scales, from yearly down to hourly time scales. The lysimeter values $E_L$ and $E_{L0}$ are used as reference. The analysis is based on the period 1 June 2009 until 31 December 2015, respectively on the six hydrological years therein (i.e., 2009/2010 to 2014/2015). For consistency with $E_L$ (see Section 2.2), also only the upward fluxes are considered for $E_{EC}$ (i.e., thus based on (neutral to) unstable conditions only). This may result in sums of $E_{EC_H}$ to become higher than the sums of $E_{EC_BOWEN}$ (depending on the distribution of the negative hourly fluxes, see Figures 5, 6). Note that the absolute sum of negative $E_{EC}$ amounts to 2.3% of positive $E_{EC}$. 

Deleted: 20

Deleted: 20
Table 2 summarizes the evapotranspiration values of the different methods for the hydrological years 2009/10 to 2014/15. The lysimeter estimates \( (E_L) \) range between 537 mm \( \text{yr}^{-1} \) and 704 mm \( \text{yr}^{-1} \). Setting evapotranspiration to zero during precipitation events \( (E_{L0}) \) allows for a comparison with the estimates of the eddy covariance method and reduces the values remarkably to a range of 521 mm \( \text{yr}^{-1} \) to 672 mm \( \text{yr}^{-1} \) (by up to -8% and -5% on the average). Except for \( E_{EC,E} \), the eddy covariance estimates show mostly lower values than the lysimeter estimates.

The monthly evolution of \( E_{L0} \) (and \( E_L \) for comparison) displayed in Figure 5 displays a pronounced seasonal cycle with highest values in summer and lowest values in winter. \( E_{L0} \) in spring is higher than in autumn. The difference of the monthly EC estimates to \( E_{L0} \) (Figures 5, 6) exhibits a seasonal cycle as well, with highest absolute differences in summer when the highest fluxes occur and highest relative differences in winter. The EC estimates mostly underestimate (overestimate) \( E_{L0} \) in summer (spring), while there is no clear tendency during autumn and winter. The \( E_{EC,BOWEN} \) estimates based on the daily force-closure (see Section 2.3) show a consistent temporal evolution but a reduction of EC evapotranspiration compared to the one based on the hourly force-closure (see Supplementary Figure 5). The underestimation of the EC estimates in summer based on the hourly force-closure thus becomes slightly larger, while the rather positive EC bias in winter turns into a predominantly negative bias.

A similar picture results from the analysis of daily (not shown) and hourly evapotranspiration values. Regarding the latter, Figure 7 displays a scatterplot of hourly values from our reference lysimeter-based evapotranspiration measurement \( E_{L0} \) and from the different EC evapotranspiration estimates for the raw measured data (excluding gaps; Figure 7). Overall, all EC-based estimates appear to underestimate \( E_{L0} \) (slopes of less than 1, and negative biases except for \( E_{EC,B} \)), in particular for high values. However, the \( R^2 \) values are similar for the different EC estimates. In addition, hourly \( E_{EC,BOWEN} \) and \( E_{L0} \) is compared for different meteorological conditions and times of the day (see Supplementary Tables 1-3). 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 (Supplementary Tables 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 (Supplementary Table 1).
4 Discussion

4.1 Temporal/spatial representativeness and differences between measurements

The analysis of the hydrological year-based evapotranspiration values derived from the catchment water balance approach $E_C$ and from the lysimeter $E_L$ (Figure 3, Table 1) shows that the considered period with parallel EC and lysimeter measurements is representative and well covering the climatological variability at the site (see also Figure 1). Despite the scale discrepancy (catchment vs. local scale), $E_C$ and $E_L$ overall show a similar magnitude of the fluxes. In addition, a previous comparison of the lysimeter seepage and whole-catchment runoff at monthly time scale revealed a high correlation in these measurements (Seneviratne et al., 2012). Based on this, we infer that the local-scale lysimeter measurements are representative for the whole catchment.

The difference between the EC evapotranspiration estimates and the lysimeter evapotranspiration $E_{L0}$ displays a seasonal dependency with highest absolute differences in summer, with $E_{EC, Bowen}$ showing about 10% lower values during this season (on the monthly time scale). On the other hand, highest relative differences occur in winter (as shown in Figures 5, 6). Despite these differences, it is important to note that overall good agreement between the two measurement techniques are achieved.

4.2 Limitations/errors in the measurements

In terms of instrumental uncertainty of EC measurements, attention was spent on the surface heating effect of the IRGA and its influence on the measured fluxes (e.g., Burba et al., 2008). Reverter et al. (2011) showed that for evaporation fluxes the impact is much smaller than for CO$_2$. The instrument at the site is tilted downwards at an angle of 45° in order to minimize the impact on the near-infrared signal by direct solar radiation, as it faces south, and thus to minimize the surface heating effect. This orientation of the IRGA also reduces the accumulation of rainwater on the optical element (see Section 2.3).

Recently, the accuracy of the vertical wind component $w$ measured with non-orthogonal sonics and the temperature measured by sonics have been under discussion as a potential source of lack of closure. But the different studies disagree on the impact (Loescher et al., 2005; Mauder et al., 2007; Burns et al., 2012; Frank et al., 2013; Kochendorfer et al., 2012; Mauder, 2013), and further investigations are thus needed.

In addition, a drawback of the EC method is the commonly observed closure gap of the energy balance (see Section 2.3). The possible underlying reasons are discussed extensively in literature (e.g., Mahrt, 1998; Foken, 2008; Franssen et al., 2010, Aubinet et al., 2012; Leuning et al., 2012). Similar values and daily patterns of the energy balance closure are reported in numerous studies (e.g., Wilson et al., 2002; Franssen et al., 2010) and are also consistent with the data presented in this study. The complex determination of evapotranspiration with the EC method does not allow for a simple error analysis. The
closure gap of the energy balance is taken here as an appropriate error estimate assuming an error of less than 5% in the measurement of the available energy. Thus, the bias in the turbulent fluxes is assumed to be of the order of 5-15% with a clear daily cycle and higher relative (but smaller absolute) errors during nighttime.

Our analysis also points at another contributor to an overall underestimation of turbulent fluxes by the EC method, namely the lack of reliable latent heat flux measurements during and following precipitation events. Because of its sensitivity to precipitation, the operation of an open-path IRGA at a rainy site results in a larger fraction of erroneous data in the $E_{EC}$ time series. The assumption is made that $E_{EC}$ is zero during hours with precipitation (see Section 2.3). But the difference between $E_L$ and $E_{EC}$, which reaches up to 8% (with an underestimation for $E_{EC}$), shows that this assumption can lead to substantial underestimation of actual evapotranspiration, e.g., because it does not always rain during the entire integration period and it takes a while after a precipitation event until the optical elements are dry again. This is an interesting result for the interpretation of the EC measurements, as $E_{EC}$ is often found to be closer to $E_{EC}$ than to $E_L$ in this study.

Figure 8 shows the energy balance closure analysis as calculated by masking precipitation hours (left part) compared with the same analysis taking into account all hours (i.e., including precipitation hours, right part). As mentioned in Section 3.2, the energy closure amounts to 86.4% (for daytime and masked precipitation) for the measured turbulent fluxes (i.e., no gap correction). For the three applied force-closing methods (see Section 2.3) the gap becomes closed per definition. When precipitation hours are included in the analysis, the EC-based energy balance gap becomes larger due to the additional contribution of available energy during these hours (i.e., 81.5% closure). Applying a correction for the missed latent heat flux during precipitation hours (based on the $E_L$ vs. $E_{EC}$ comparison of Table 2, i.e., increasing latent heat flux by 5% on the average) results in a closure of 84.4%. This shows that the amount of underestimation of latent energy during precipitation hours is substantial compared to the overall energy imbalance, contributing about 15% to the uncorrected gap (cf. 81.5% vs. 84.4%). Applying in addition the energy gap correction of the precipitation-masked data (i.e., from Figure 8, left part) results in closures of 97.1%, 97.5%, and 97.7%, respectively. Thus, the additional amount of evapotranspiration during precipitation hours as estimated by the lysimeter corresponds well with the independent measurements of available energy during these periods. It should be noted here that the amount of underestimation of evapotranspiration during precipitation periods scales with the time step used in equations (1) and (2).

The site-specific error in the lysimeter evapotranspiration is discussed in Seneviratne et al. (2012). An overall measurement uncertainty of 5-10% is assumed, whereas higher errors (20% or more, e.g., Gurtz et al., 2003) can be assumed during periods of snow cover. In the latter case, also errors in the precipitation measurement (high undercatch under snow fall, e.g., Rasmussen et al., 2012) strongly contribute to errors in $E_L$. While these errors are large in relative terms during the affected months, they are nonetheless only affecting $E_L$ when it is at very low values (late fall, winter, and early spring; see Figure 1).
Hence, they should not substantially affect the lysimeter evapotranspiration during the growing season, nor when aggregated on the yearly time scale.

5 Summary and Conclusions

We examined and compared two well-established methods to measure evapotranspiration at the site level: lysimeter-based measurements ($E_L$), which are common in hydrological research, and eddy-covariance flux measurements ($E_{EC}$), which are widely used in micrometeorology. For the analyses, we employ parallel measurements based on these two methods carried out at a research catchment in northeastern Switzerland and covering the time period 1 June 2009 to 31 December 2015. Over this multi-year time period, the measurements were compared on yearly down to hourly time scales. Moreover, they are related to a 40-year long lysimeter evapotranspiration time series.

Overall, the lysimeter and EC measurements agree well, in particular on the annual time scale. Also, the long-term lysimeter evapotranspiration agrees well with a catchment-wide estimate of evapotranspiration based on the catchment water balance over hydrological years (and assuming no changes in storage). This emphasizes the representativeness of the site-level lysimeter and EC measurements for the entire catchment despite their comparatively small source areas. We note, however, that the agreement is closest when the two time series are processed in the same manner, i.e., setting hourly evapotranspiration to zero during hours with precipitation ($E_{L0}$ for lysimeter record). The lysimeter measurements actually reveal the occurrence of non-negligible evapotranspiration fluxes during these periods, which leads to an underestimation of 5% on average of the total evapotranspiration fluxes with this processing. Hence, the lack of reliable EC measurements from the open-path IRGA immediately following precipitation events significantly contributes to the overall underestimation of latent heat flux from EC measurements, at least for humid sites such as Rietholzbach. Given this issue of underestimation of the EC evapotranspiration in hours with precipitation, a correction based on lysimeter estimates for these specific time periods could be possibly envisaged in future studies for humid sites, in addition to the correction for the energy balance closure gap. We further note that the difference between the EC and $E_{L0}$ lysimeter evapotranspiration shows a seasonal cycle, but the same pattern on different time scales (monthly to hourly).

In conclusion, our results still highlight remaining uncertainties in the various methods and techniques available to measure and estimate evapotranspiration. Nonetheless, the good agreement of the different methodologies on yearly time scale is an important result in the context of long-term water balance studies. In addition, our results emphasize the value of parallel lysimeter- and EC-based measurements to characterize the respective errors of these measurement systems.
Author contribution

I.L. implemented the initial analyses for this study and co-wrote a first version of the manuscript with S.I.S. M.H. and D.M. revised and extended the evaluation and revised the manuscript together with S.I.S. S.I.S. initiated and oversaw the project.

Acknowledgements

We acknowledge the Federal Office for the Environment (FOEN/BAFU) for providing the runoff data, MeteoSwiss for data, as well as ETH Zurich for funding. We also thank Karl Schroff for his technical support in setting up and maintaining the measurement site.
565 References


Jaun, S., 2003. Evapotranspiration und Strahlungskomponenten im Forschungsgebiet Rietholzbach [Evapotranspiration and radiation components in the Rietholzbach research catchment] [in German], MSc Thesis, Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland.


Peter, I., 2011. Turbulence measurements and footprint estimates at Rietholzbach. MSc Thesis, Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland.


Figure 1: Box plots showing the median, interquartile range as well as minimum and maximum of monthly climatological values (in black, climatological mean included as black square) with the monthly values for the period of investigation 2009-2015 (in colour) for: (a) air temperature $T_{\text{air}}$, (b) precipitation $P$, (c) net radiation $R_n$, (d) catchment runoff $Q_C$, (e) lysimeter evapotranspiration $E_L$, and (f) lysimeter seepage $Q_L$. 
Figure 2: (left) Aerial map of the site (see https://s.geo.admin.ch/6de2decf3b5 for an interactive online version of the map) and (right) schematics of the EC tower (denoted T), lysimeter (L) and radiation (R) measurement setup as well as the frequency of wind direction and velocity at the site “Büel” (white rectangle defines the area of the measurement field). The distance between the tower and the sonic (S) volume equals to 1.17 m. The wind sector obstructed by the tower and the IRGA is highlighted in red and masked for the analyses (from 310 to 50 degree). The identical hatching of the surrounding grassland and the lysimeter surface indicates that the latter is treated according to the former (see text for details).
Figure 3: Catchment water balance for the hydrological years (i.e., October-September) 1976/77 until 2014/15. $E_c$ denotes catchment evapotranspiration, $E_L$ lysimeter evapotranspiration, $Q_C$ catchment runoff, $Q_L$ lysimeter seepage, and $P$ precipitation (corrected according to Gurtz et al. 2003, Table 1 therein). The lines show the ratio $E_c/P$ (dark grey) resp. $E_L/P$ (blue).
Figure 4: (a) Sum of turbulent fluxes (i.e., sum of sensible heat flux $H$ and latent heat flux $\lambda E$; $H+\lambda E$) versus the available energy (i.e., net radiation $R_n$ minus surface soil heat flux $G$; $R_n-G$) and (b) mean daily pattern of the energy balance components. Graphs are based on measured hourly values (i.e., excluding gap-filled data, and masked for precipitation and wind directions affected by the tower) for the time period 1 June 2009–31 December 2015.

Comment [41]: Updated to include measured values only (i.e., excluding gap-filled data)

Deleted: hourly values
Deleted: and masked for precipitation and wind directions affected by the tower
Figure 5: Monthly values of the different evapotranspiration estimates (with $E_L$ denoting lysimeter evapotranspiration, $E_{L0}$ lysimeter evapotranspiration with values set to zero during hours with precipitation, and $E_{EC_{BOWEN}}$ EC-based evapotranspiration corrected according to the Bowen ratio) for the time period June 2009 to December 2015. The black bars indicate the range based on $E_{EC_H}$ and $E_{EC_E}$ (i.e., $E_{EC}$ corrected by assigning the energy balance closure gap to sensible heat flux only and to latent heat flux only, see Section 2.3). Note that from July to October 2014 an energy gap correction is not possible due to missing soil heat flux (see Section 2.5) and thus $E_{EC_{BOWEN}}$ and $E_{EC_E}$ are not available.
Figure 6: Monthly relative differences between lysimeter evapotranspiration \( E_{L0} \) and EC-based evapotranspiration \( E_{EC} \), i.e., \( \frac{E_{EC} - E_{L0}}{E_{L0}} \) versus the absolute values of \( E_{L0} \). Different seasons are displayed in different colours. The points indicate \( E_{EC, Bowen} \) (corrected according to the Bowen ratio) and the black bars indicate the range based on \( E_{EC, H} \) and \( E_{EC, E} \) (\( E_{EC} \) corrected by assigning the energy balance closure gap to sensible heat flux only and to latent heat flux only). Note that the July to October 2014 values with missing \( E_{EC, Bowen} \) and \( E_{EC, E} \) (see Section 2.5) are omitted.
Figure 7: Comparison of hourly EC-based evapotranspiration $E_{EC}$ with lysimeter evapotranspiration $E_{L0}$ based on measured values (i.e., excluding gap-filled data and masked for precipitation and wind directions affected by the tower) in the time period 1 June 2009–31 December 2015 (n=30615 for $E_{EC_H}$ respectively n=30002 for $E_{EC_BOWEN}$ and $E_{EC_E}$). The comparison is shown separately for $E_{EC}$ corrected according to the Bowen ratio ($E_{EC_BOWEN}$) and $E_{EC}$ corrected by assigning the energy balance closure gap to sensible heat flux only ($E_{EC_H}$) and to latent heat flux only ($E_{EC_E}$).

Comment [42]: Updated to only include measured values (i.e., excluding gap-filled data) both from EC and from the lysimeter.

Deleted: estimates
Deleted: s
Deleted: n=39525
Figure 8: Yearly aggregated available energy ($R_n - G$) vs. sum of turbulent fluxes (for daytime, time period June 2009 to December 2015, excluding gap-filled data), with percentages denoting the amount of closure. (Left part) Totals calculated from hourly data masked for precipitation. The energy closure amounts to 86.4% for the measured turbulent fluxes (i.e., no gap correction) and the gap becomes per-definition closed for the three applied corrections (i.e., correction according to the Bowen ratio $EC_{BOWEN}$, and correction by assigning the energy balance closure gap to sensible heat flux only $EC_{H}$ and to latent heat flux only $EC_{E}$, see Section 2.3). (Right part) Totals calculated by including also precipitation hours. Here the gap is corrected by applying a correction for missed evapotranspiration during hours with precipitation (based on the lysimeter evapotranspiration estimates $E_L$ and $E_{L0}$; denoted $E_L$ vs. $E_{L0}$ correction) plus considering the energy gap correction based on the precipitation-masked data (see text for details).
Table 1: Statistical properties of the catchment water balance for the hydrological years (i.e., October-September) 1976/77 until 2014/15 and the absolute values for the hydrological years 2009/10 to 2014/15. \( P \) denotes precipitation, \( Q_c \) catchment runoff, \( Q_L \) lysimeter seepage, \( E_c \) catchment evapotranspiration, and \( E_L \) lysimeter evapotranspiration. Units in mm yr\(^{-1}\) (except for \( E_c/P \) and \( E_L/P \), which are dimensionless).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>mean</td>
<td>1598</td>
<td>1591</td>
<td>1281</td>
<td>1967</td>
<td>193</td>
<td>1459</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>1591</td>
<td>1591</td>
<td>1281</td>
<td>1967</td>
<td>193</td>
<td>1459</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>1446</td>
<td>1177</td>
<td>1764</td>
<td>168</td>
<td>(168)</td>
<td>(1715)</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>1940</td>
<td>2031</td>
<td>2016</td>
<td>1605</td>
<td>1151</td>
<td>716</td>
</tr>
<tr>
<td></td>
<td>standard deviation</td>
<td>468</td>
<td>751</td>
<td>852</td>
<td>204</td>
<td>293</td>
<td>325</td>
</tr>
<tr>
<td>1976-2007 climatology</td>
<td>1281</td>
<td>101</td>
<td>901</td>
<td>172</td>
<td>362</td>
<td>526</td>
<td>617</td>
</tr>
<tr>
<td>( Q_c )</td>
<td>mean</td>
<td>1006</td>
<td>967</td>
<td>716</td>
<td>1471</td>
<td>185</td>
<td>1063</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>987</td>
<td>967</td>
<td>716</td>
<td>1471</td>
<td>185</td>
<td>1063</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>906</td>
<td>716</td>
<td>1471</td>
<td>185</td>
<td>1063</td>
<td>1151</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>1026</td>
<td>1070</td>
<td>187</td>
<td>-</td>
<td>1140</td>
<td>675</td>
</tr>
<tr>
<td>( E_c )</td>
<td>mean</td>
<td>592</td>
<td>584</td>
<td>364</td>
<td>880</td>
<td>101</td>
<td>396</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>592</td>
<td>584</td>
<td>364</td>
<td>880</td>
<td>101</td>
<td>396</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>507</td>
<td>316</td>
<td>864</td>
<td>101</td>
<td>396</td>
<td>679</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>618</td>
<td>675</td>
<td>1245</td>
<td>881</td>
<td>864</td>
<td>1140</td>
</tr>
<tr>
<td>( E_c/P )</td>
<td>mean</td>
<td>0.37</td>
<td>0.37</td>
<td>0.23</td>
<td>0.50</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>0.37</td>
<td>0.37</td>
<td>0.23</td>
<td>0.50</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>0.34</td>
<td>0.26</td>
<td>0.49</td>
<td>0.06</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>0.44</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>( E_L )</td>
<td>mean</td>
<td>641</td>
<td>566</td>
<td>433</td>
<td>704</td>
<td>55</td>
<td>560</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>641</td>
<td>566</td>
<td>433</td>
<td>704</td>
<td>55</td>
<td>560</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>557</td>
<td>352</td>
<td>704</td>
<td>55</td>
<td>560</td>
<td>589</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>667</td>
<td>704</td>
<td>704</td>
<td>704</td>
<td>542</td>
<td>537</td>
</tr>
<tr>
<td>( E_L/P )</td>
<td>mean</td>
<td>0.36</td>
<td>0.36</td>
<td>0.26</td>
<td>0.49</td>
<td>0.06</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>0.36</td>
<td>0.36</td>
<td>0.26</td>
<td>0.49</td>
<td>0.06</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>0.34</td>
<td>0.28</td>
<td>0.52</td>
<td>0.06</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>0.49</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.49</td>
<td>0.49</td>
</tr>
</tbody>
</table>

\(^1\)from Seneviratne et al. (2012), based on calendar-year (January-December) values. For \( P \) (and \( E_L/P \)) the statistics based on undercatch-corrected and uncorrected (in brackets) \( P \) are provided (see Section 2.4).
Table 2: Lysimeter ($E_L$ and $E_{LB}$) and EC ($E_{EC}$) evapotranspiration (including $E_{EC}$ corrected according to the Bowen ratio ($E_{EC\_BOWEN}$), and $E_{EC}$ corrected by assigning the energy balance closure gap to sensible heat flux only ($E_{EC\_H}$) and to latent heat flux only ($E_{EC\_E}$)) for six hydrological years and the respective six-year averages. Percentages denote the differences of $E_{EC}$ and $E_L$ to $E_{LB}$. Note that for 2013/14 and 2014/15 an energy gap correction is not possible for a four-month period due to missing soil heat flux (see Section 2.5) and thus $E_{EC\_BOWEN}$ and $E_{EC\_E}$ are not available. Units in mm yr$^{-1}$.

<table>
<thead>
<tr>
<th>Hydrological year</th>
<th>$E_L$</th>
<th>$E_{LB}$</th>
<th>$E_{EC_BOWEN}$</th>
<th>$E_{EC_H}$</th>
<th>$E_{EC_E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/10</td>
<td>589 (+8%)</td>
<td>543</td>
<td>526 (-3%)</td>
<td>473 (-13%)</td>
<td>555 (+2%)</td>
</tr>
<tr>
<td>2010/11</td>
<td>704 (+7%)</td>
<td>659</td>
<td>614 (-7%)</td>
<td>545 (-17%)</td>
<td>652 (-1%)</td>
</tr>
<tr>
<td>2011/12</td>
<td>704 (+5%)</td>
<td>672</td>
<td>622 (-7%)</td>
<td>544 (-19%)</td>
<td>663 (-1%)</td>
</tr>
<tr>
<td>2012/13</td>
<td>542 (+4%)</td>
<td>521</td>
<td>547 (+5%)</td>
<td>476 (-9%)</td>
<td>589 (+13%)</td>
</tr>
<tr>
<td>2013/14</td>
<td>537 (+2%)</td>
<td>526</td>
<td>NA</td>
<td>506 (-4%)</td>
<td>NA</td>
</tr>
<tr>
<td>2014/15</td>
<td>647 (+1%)</td>
<td>638</td>
<td>NA</td>
<td>547 (-14%)</td>
<td>NA</td>
</tr>
<tr>
<td>Average</td>
<td>621 (+5%)</td>
<td>593</td>
<td>577 (-3%)</td>
<td>515 (-13%)</td>
<td>615 (+4%)</td>
</tr>
</tbody>
</table>