We thank the reviewers for their positive feedback and valuable suggestions for improving the manuscripts. We have carefully revised the manuscript according to the constructive and helpful comments. For clarity and visual distinction, all responses have been marked in blue. We hope the manuscript is now considered to be suitable for publication in AMT.

[0] This could potentially be an interesting paper comparing long-term water vapor flux measurements using open-path (OP) and closed-path (CP) systems above a temperate forest in China. The novelty of the study could be better emphasized but it is true that there are much fewer long-term studies comparing OP and CP water vapor fluxes than it is the case for CO2 fluxes. The vegetative region in China is underrepresented in these measurements (e.g. evapotranspiration). Long-term measurements using OP and CP systems can be helpful to characterize and reduce the uncertainties in each season in which changes in relative humidity and other environmental parameters can impact the accuracy of fluxes measured by OP and CP systems.
Response: we thank the referee for his/her positive comments.

General
Although the paper makes a promising first impression, after reading it more thoroughly it becomes apparent that significant work is needed to meet the AMT standards. The story is unfocused and generally unclear which often misleads the reader. It promised long-term measurements but the actual comparative fluxes are demonstrated only for five days (Figure 3 and 4). The annual data in Fig 5 therefore does not seem to be justified. A more extensive period of calculated fluxes including gapfilling (e.g. highlighted with different color) would be more convincing.
Response: for clarity, only a short-term and typical daily course of LE was presented in Fig.3 and Fig.4 (now Fig.4 and Fig.5). The gapfilled data were not presented in Fig 5a (now Fig.6a) is because that it has nothing to do with the intercomparison of OP and CP system.

It is paramount to observe quality control protocols and account for all the major factors behind the accurate measurements. The authors point to the importance of accurate lag-time determinations but it is unclear which dataset was processed with which lag time approach. This is unfortunate because demonstrating the importance of lag time could have convinced those from the community who regard the lag-time component as trivial.
Response: thanks for reminding us this issue. We have added such information in the revised manuscript.

Figure 3 shows counter-intuitively that the datasets are not much different regardless of which lag-time approach is used. This is very surprising because the authors discuss that the lag time was not constant and therefore inaccurate lag-time would be expected to cause significant underestimation (e.g. Taipale et al., 2010).
Response: we don't think there is any contradiction between our results and Taipale's findings. Taipale et al (2010) presented a straightforward comparison of five lag time methods to assess their applicability to DEC measurements with PTR-MS, and also found that the constant lag time methods had a tendency to underestimate the absolute values of fluxes, whereas the maximum

covariance method was prone to overestimation.
Because the constant lag time leads to systematic underestimation of LE, so despite the visual difference in Fig.3 (now Fig.4) is not remarkable, it do yield statistically significant underestimation.

It is suggested that the authors make a deeper insight into the interpretation of the results, clarify in the methods how exactly (step by step) each flux dataset was treated and work on the overall coherence of the story. In addition the comments below should be addressed before the publication in AMT.
Response: we have carefully revised the manuscript according to the constructive and helpful comments. More detail please see our response to comment below.

Major
1) The authors made a good effort to describe corrections applied to the fluxes. While they describe the corrections it is not clear how and to which datasets they have been applied. It could be useful for a reader to have this information summarized in a table (e.g. expanded Table 1 or as a new table) so that it is clear which corrections were included to the CP and OP systems and if there were any differences in how the OP and CP datasets were treated.
Response: we have further detailed the description of each correction result for both OP and CP datasets in the revised manuscript, while the table was not presented considering that actually there is no so much information for a table.

2) It is mentioned that 15.7 % of data were rejected due to nonstationarity. Were the fluxes from the second class (i.e. low quality) permitted? It would be useful to know how much data were in each quality category for each correction. It is also unclear if the authors detrended the datasets before the flux calculation and if the night time data below the u* threshold were rejected.
Response: thanks for these details. We classified the data quality with 3 classes. The data of classes 1–2 were used in this study, and class 3 were excluded as abnormal.
The datasets were linear detrended before axis rotation for tilt correction rotation.
The u* filtering was not applied to the datasets. We don't think that water vapor, as CO2, could be accumulated below canopy at low u* value, considering that the molecular weight of water vapor is far less than that of air.
We also clearly stated these in the revised manuscript.

3) Water vapor fluxes can be underestimated for a number of reasons, including mismatched fetch, inhomogeneous surface characteristics, lack of coincidence of source areas (e.g. leaves, soil) (e.g. Mahrt, 1998). The authors do not discuss here moisture advection or flux divergence. The question is if the impressive agreement between the LE fluxes measured by CP and OP can be for the right or wrong reason? The discussion should be expanded to include all important factors.
Response: we agree with the reviewer that several other factors could also cause the underestimation of LE measured from eddy covariance method. However, all the factors mentioned above have the same impacts on both OP- and CP system considering that they

shared one sonic anemometer. Therefore the Discussion section is only focus on the possible reasons that could cause discrepancy between OP- and CP system.

4) The major problem with the data is clearly seen in Figure 1 which I have pointed to during access review to give authors a chance to clarify these discrepancies. P4718 L11 "As shown in Fig. 1, the fluctuations in water vapor density recorded from the OP analyser are similar to those from the CP analyser, and both follow the same fluctuation trends." It is extremely surprising how these two traces can look similar to the authors. Further the authors write: "Overall, the magnitude of fluctuations in water vapor densities measured from OP system is approximately 1.2 times higher than that from the CP system." –The amplitudes seem to me at least a factor of two different. Iam therefore concerned about the quality of the presented data. I had thought that the authors showed different periods by mistake, but looks like there must be a problem with the data if these are the same periods (after lag-time correction). Do the authors suggest that the tube attenuation could completely change the signal? This reviewer is not convinced.

Response: because the spatial separation of Li7500 and intake point of Li7000, as well as the attenuation of turbulence intensity in the IRGA cell, the measured water vapor concentration values are not exactly similar in magnitude and tendency, and the consistence of their fluctuation values is even worse. But statistically, the massive samples will give a similar flux values.

To better present the intercomparision, we replotted Figure 1 and updated it with a much longer dataset (30min raw data). We also added a figure to show the spectral and cospectra for the open and closed path data, and the data quality assessments were further conducted in the revised manuscript.

Specific

5) P4715 L14 Data processes or data processing?

Response: we have replaced 'Data processes' with 'data processing'.

6) P4718 L7 Why is there a correspondingly larger fraction (15.7) due to nonstationarity in CP fluxes? What was the percentage of the rejections in OP fluxes due to nonstationarity?

Response: results of integral turbulence and stationarity tests showed that this is a correspondingly larger fraction (15.7%) of rejected data from CP system compared to that of OP system. This is consistent with the findings of Haslwanter et al. 2009. Agric For Meteorol. 149(2): 291–302; and Nordbo et al (Tellus B 2012, 64, 18184). For example, the latter reported that the Flux stationarity of data measured from CP system is nearly 5% higher for that of OP system.

The possible reason for this is that the stationarity and intermittency criteria filter out LE data, especially during night with high relative humidity. For CP system, as we demonstrated in Figure 2, the time lags are very variable in high humidity condition, and hence calculated subperiod (5-min) fluxes, this leads to a high non-stationarity based on closed path data. Additionally, the turbulence distortion in the sample tube is also a possible reason.

7) P4718 L11 “The time series from the CP analyser was set back 9.1 s to account for the tube delay.” This is confusing. Earlier the authors wrote they used accurate lagtimes from crosscorelation which were meant to be compared with median lag-times. It should be made very clear how the lag time was accounted for in each analysis.
Response: there is no contradiction at this point. The time shift (9.1s) is only applied to the dataset in Fig.1. To avoid misunderstanding, we have moved this description to the Figure Caption.

8) P4717 L15 “Comparison of data coverage” seems to be misleading. The authors are describing here differences in data availability, but this section could fit better the methods section. A section “Flux validation and quality control” with individual CP and OP subsections could make it much clearer to a reader how the data were treated.
Response: revised as suggested. The Chapter 3.1 was moved to the METHODS section as a sub-section of “Data processing and quality control”.

9) Table 2 “For sound statistical analysis, the minimum available data for time lag analysis was 30 in each class. If the filtered data number is less than 30, its time lags are set equal to that of adjacent class.”. I am not comfortable with this statement. I think it would make sense to show the median lag time values also for the dry conditions.
Response: as suggested, we have presented the calculated median lag time values for the dry conditions.

10) P.4719 The discussion about the lag times is generally interesting but unfortunately is very unclear. It is expected that lag-times from the covariance function would not work at night due to low turbulence. Also condensation in the tube should be prevented in particular when the temperature goes down in the morning. Was the tube insulated/heated to prevent condensation? If not, the data which clearly indicate water condensation should be rejected.
Response: yes, we acknowledge that lag-times from the covariance function are irregular at night due to low turbulence, however, fortunately the actual LE values at night with low turbulence is generally very low too, it’s contribution to the daily values is negligible.
Additionally, the sample line was self regulating heated to damp out temperature fluctuations in the tube. We have added this information in the revised manuscript.

11) P.4720 L.1 “This increases the possibility that unrealistic time lags are detected and hence results in flux overestimations”. This does not make sense. If the lag time is inaccurate the flux should be underestimated not overestimated because the value would be away from the maximum peak (e.g. Taipale et al., 2010).
Response: the covariance maximum method with a broad time lag search window results in a maximal while unrealistic covariance of vertical wind speed w and H_2O, which is not always the true covariance of w and H_2O with the true time lag. So sometimes overestimation is possible. Taipale et al. (Atmos. Meas. Tech., 3, 853–862, 2010) do report that “ using a prescribed time-lag may result in a systematic underestimation of the flux as the “true” time-lag is likely to vary over time”, while it also confirmed that “The maximum covariance method was prone to overestimation of the absolute values of fluxes.”

12) P4722 L11 “But currently, there are no generally accepted standard procedures to handle the flux data.” This is very surprising to hear from the authors (see, for example, Foken et al., 2005; Mauder et al., 2008).

Response: to avoid misunderstanding, we have deleted this sentence in the revised manuscript. However we still believe that there is no generally accepted standard procedure to handle the flux data. For example, in term of title correction, there are three methods of coordinate rotation method: double rotation, triple rotation and planar fit method. Different methods would lead to somewhat different final results.

13) Conclusions are not sufficiently strong and lack clear take-home messages. In particular the last sentence is not impressive. The authors should try to summarize the novel elements of their study and highlight the key findings in response to what has been promised in the abstract.

Response: the Conclusion section has been rephrased accordingly.

14) Table 1: What do you exactly mean by "total datasets", "Technique issues", "other hard flags without no clear reason"?

Response: the term "total datasets" was removed in the revised manuscript considering that it doesn’t have much sense in this table. We have further elaborated the term "Technique issues" in the title of Table 1, and replaced the term "other hard flags without no clear reason" with "out-of-threshold values with no clear reason".

15) Table 2: if there was no data in class 1 it can be deleted. The median flux value should be used in class 2 as in other classes even though less data points is present in class 2.

Response: revised as suggested.

16) Figure 1:It is unclear why the datasets are not similar. It does not seem to be because of dampening in the tube. Perhaps the data may have been taken from completely different periods.

Response: to better present the intercomparision, we replotted Figure 1 and updated it with a much longer dataset (30min raw data).

17) Figure 3: It is recommended to replace DOY with actual date including the year.

Response: Done.

18) Figure 4: as above

Response: we would like to, but there is no much place to present the actual date in the bottom of Fig.4.

Technical

There are numerous places with spelling, typographical and language issues. The language and clarity should be significantly improved throughout. In particular some sentences can be difficult to understand for readers. Here are just a few examples: P.4712 L19-22, P. 4714

L20-21, P4718 L6-8, P4714 L15.

Response: we have tried our best to check the whole manuscript to avoid inaccurate expression and inappropriate tense of English. We hope that the content of the manuscript is clear enough to understand.

Specific:

P.4712 L19-22:

Originally the sentence read: "it also suggests that when discuss the energy balance closure problem in flux sites with closed-path eddy covariance systems, it has to be aware that some of the imbalance is possibly caused by the systematic underestimation of water vapor fluxes."

Now we changed it to: and it also suggests that some of the imbalance of the surface energy budget in fluxes sites is possibly caused by the systematic underestimation of water vapor fluxes measured with closed-path eddy covariance systems.

P. 4714 L20-21：

Originally the sentence read: "Therefore it is critical to further explore that weather the OP system measured vapor water fluxes are consistent with that of CP system"

Now we changed it to "Therefore it is critical to further examine that weather the vapor water fluxes measured with OP system are consistent with that of CP system.

P4718 L6-8:

Originally the sentence read: "There is a correspondingly larger fraction of unaccepted CP flux values too, with most of the rejected data (15.7 %) is mainly due to failing in the integral turbulence and stationarity tests.

Now we changed it to: There is a correspondingly larger fraction of unaccepted CP flux values too, with most of the rejected data is mainly due to failing in the integral turbulence and stationarity tests.

P4714 L15：

Response: we don't think there is problem with this sentence, so please more clearly point out the error.

References

Ammann, C., Brunner, A., Spirig, C., Neftel, A. (2006). Technical note: Water vapour concentration and flux measurements with PTR-MS. Atmospheric Chemistry and Physics, 6(12), 4643-4651.

Foken, Thomas, Mathias Göockede, Matthias Mauder, Larry Mahrt, Brian Amiro, and William Munger. "Post-field data quality control." In Handbook of micrometeorology, pp. 181-208. Springer Netherlands, 2005.

Mahrt, L. (1998). Flux sampling errors for aircraft and towers. Journal of Atmospheric and Oceanic technology, 15(2), 416-429.

Mauder, M., T. Foken, R. Clement, J. A. Elbers, W. Eugster, T. Grünwald, B. Heusinkveld, and O. Kolle. "Quality control of CarboEurope flux data? Part 2: Intercomparison of eddy-covariance software." Biogeosciences 5, no. 2 (2008): 451-462.

Taipale, Risto, Taina M. Ruuskanen, and Janne Rinne. "Lag time determination in DEC measurements with PTR-MS." Atmospheric Measurement Techniques3, no. 4 (2010): 853-862.