Answers to Referees (Comments from the Referees are in black, answers in blue)

Referee 2

1) The title of this paper is inappropriate. Although the paper does provide a basic description of the EddyUH software, the primary purpose of the paper appears to be a comparison of the effect of different processing algorithms on fluxes from two example data sets.

We agree with the reviewer, and we decided to change the title as:

Quantifying the uncertainty of eddy covariance fluxes due to the use of different software packages and combinations of processing steps for a wide range of instrumentation in two contrasting ecosystems

2) The authors undoubtedly spent much effort to achieve the comparisons presented in this paper. However, whether the stated goal of "estimate(ing) the flux uncertainty due to the use of different software packages" (page 1 line16), and whether it is true that the "processing steps were consistent between the software packages" (page 1 line 21) is questionable. Instead what the authors present is a comparison of different algorithms, with an inevitable difference in the calculated fluxes. This problem appears to stem primarily from the frequency response correction algorithms. There is no reason to believe that different algorithms for correcting frequency responses should give the same result, any more than we expect different coordinate rotation algorithms to result in the same velocity vectors. The best a software author can do is to provide access to as many algorithm versions as is practical, and to expose as many algorithms configuration options to the user – as is practical. However, if and where identical algorithms are present in the two compared softwares, then a true software comparison should be made, if only to ensure that a near identical result is obtained. This should extend beyond basic checking algorithm form as there may be other embedded assumptions in the software such as a molecular weight of methane may have a value of 16.00 in one software and 16.04 in another software, or one software may use single precision calculations while another uses double precision. Such situations may lead to differences in the results obtained from different software packages. This is perhaps what the authors refer to as "tuning" (page 2 line 22) and is actually quite important for the purpose of identifying software bugs, though perhaps not all that interesting when trying to publish a paper.

As an answer for this comment, we would like to point out that on page 1 line 21 we are not saying that the processing steps were equal, but that they were consistent. We do not want to get exactly the same results, but the goal is to use the optimal configuration of the two software packages, and estimate the uncertainty. Indeed, the goal of this study was not to identify software bugs, which we assume have been sufficiently debugged from the software packages widely used by the community.

3) The authors also seem to imply that fluxes of CH4 and N2O inherently different from those of H2O and CO2, from a software perspective. They are all trace gas fluxes, and apart from sensor specific corrections, should be subject to the same processing stream. However, the poorer frequency response, or sampling path characteristics of these sensors may expose weaknesses in the algorithms applied for correcting these fluxes.

We do agree with this comment, especially in relation to the fact some steps/corrections require some level of optimization for CH4 and N2O, which often show low signal to noise ratio and episodic fluxes, For instance, Felber et al (2015) showed that despiking of CH4 time series can be difficult due to a highly variable CH4 signal and thus methods developed for CO2 are not necessarily directly applicable to CH4 time series.

4) In section 2.1 the authors give a one page description of the software, which seems rather short and procedural considering that the software should be the main focus of the article. There is no discussion as to why it was designed in the way it was, or why its design may be an improvement over other software designs.

We have now removed this section, and moved the EddyUH specific description to a new Appendix A, while a short introduction to the two software is now combined in the new chapter 2.3 *Software description and setup of software runs*. Some text related to the point: "why it was designed in the way it was, or why its design may be an improvement over other software designs" was added in the chapter.

5) In section 2.4 a step-wise assessment is taken for comparing the effect of the algorithms in the two softwares by removing key correction steps. This is fine, but there appears to be no effort to account for the differences. One would assume that to compare the effect of an algorithm in the processing chain that all preceding a subsequent algorithms are behaving identically between the two software packages.

We do agree with the Reviewer. However, we have used identical processing steps as much as possible. Some differences in algorithms were inevitable due to different implementations of algorithms in the softwares i.e. identical algorithms were not available for all processing steps. Finally, these differences are anyway fully described and discussed in the manuscript.

6) On page 7 line 6 the authors give the 'best' agreement between EddyUH and Eddypro, and then list five flux comparisons. Surely the best comparison will consist of a single flux comparison – unless by chance all five flux comparisons had exactly the same statistical result.

We agree and we have replaced "best" with "very good".

7) Page 7 line 9: did you instead mean "…no significant systematic differences…"

Yes, we have changed it.

8) On page 7 line 12 you attribute the scatter to spurious, unrealistic spectral correction values. On what justification can you say that the corrections are unrealistic because they are more than _ 50%. While such large corrections may not be desirable, they certainly do occur and should be accounted for.

Yes, we think that these few values with high spectral corrections are not real, since the flux measurements were done using open-path LI-7500 located very close to the sonic anemometer and 23 meters above the ground and with such setup high signal attenuation is unrealistic. Thus we attributed these high spectral corrections to the fact that the method implemented in EddyPro utilize the measured 30 min WT cospectra (and not the cospectral model) to calculate the spectral correction factors (Fratini et al., 2012). Although default quality criteria are applied for selecting good spectra, there may be noised cospectra, which may give highly uncertain values of the correction factors for single half-hour periods. We have now noticed that, since the version 6.0.0 of EddyPro released on 01.08.2015, the criteria for cospectra filtering are explicitly set by the user, giving the possibility to eliminate these periods with unrealistic spectral correction values.

9) On page 7 line 29 you indicate a difference in the WPL corrections for the latent heat fluxes. Can the authors explain why the WPL correction should differ? If it is because the inputs to the WPL correction differ then that is another matter and attributable to a different processing step. If, however, the inputs are the same and still the WPL correction differs then you need to check your code for mistakes.

Thanks for pointing out this. We have now checked again, and found the reason for this difference, which is due to a difference in one of the input of the WPL correction. In fact, the water vapour densities (ρ_w) used in the WPL T-term are different between the softwares. The reason is that EddyUH uses the one from LI-7500, while EddyPro calculates ρ_w from meteo RH data. As seen in the figure below, if we force to use the same ρ_w the difference in the WPL T-term disappears. We will add one sentence related to this point in the revised manuscript.
This also explains why the same difference after WPL correction cannot be seen in CO2 flux measured by LI-7500 (see Fig. 4j in the manuscript), where in the WPL T-term ρ_w is replaced by CO2 concentration density (ρ_c). (Note that ρ_w still appears in the factor $\mu\sigma$, where μ is the ratio of molar masses of dry air and water vapour, and σ the ratio of the densities of water vapor and dry air. However, this factor is very small compared to the other terms, and the effect negligible.)

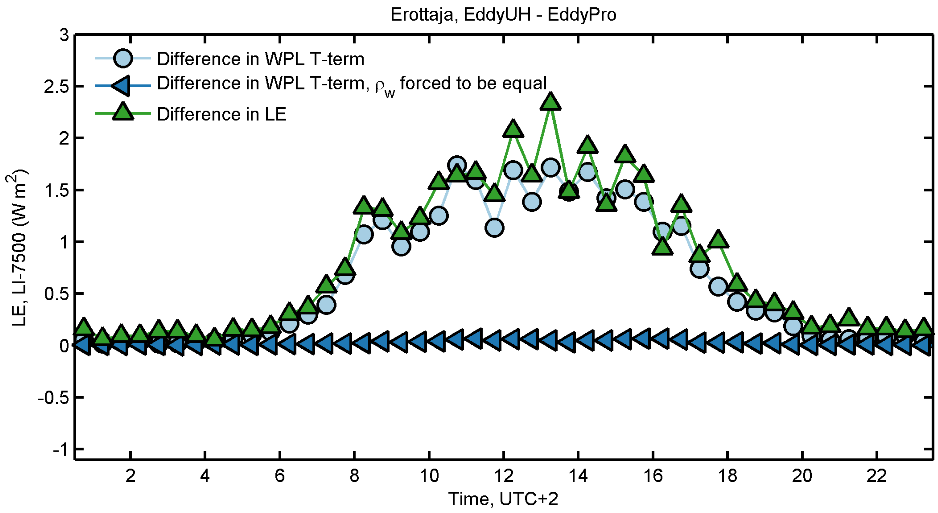

Figure 1. Green triangles are the difference in LE LI-7500 between EddyUH and EddyPro (after WPL correction, equals the "WPL" curve plotted in Fig. 4r in the manuscript) plotted together with the difference in WPL T-term (circles). The small differences between green triangles and circles are due to small differences in the factor $\mu\sigma$. If we force the ρ_w to be the same in the two softwares, the difference in WPL T-term disappears (blue triangles).

10) Similar to the last comment, the difference in humidity dependent lag times (Page 11 line 22) is quite interesting and potentially quite important but little explanation as to why there is a difference. Is it because the input data differ, or are there some presumable identical processes in the two softwares differing in some way.

We guess the reviewer is referring here to the response times (and not to the lag times). Differences come from different ways the two software estimate the response times. In the revised manuscript we will add some discussion related to this point.

11) On page 12 line 3 the authors attribute wet surface conditions as causing a large WPL water vapour term correction on the CO2 fluxes. Wet surfaces do not cause a WPL correction, only water vapour fluxes.

Yes, we agree and we reformulated the sentence as: *".....due to large H_2O fluxes (average daytime value of LE equals 170 W m^{-2}) caused by the wet surface conditions and the presence of vegetation at the site."*.

12) The terms density correction, WPL correction, and dilution correction all appear in the paper. It might be best to just refer to the WPL correction as it encompasses both the dilution effect of evapotranspiration and the ideal gas law heating related volume effects. Further, I suggest the authors refer to a band-broadening correction instead of using the rather indistinct "spectroscopic correction"; which refers to any correction to an optically based measurement.

Yes, we partly agree and we will mainly use the term "WPL correction" in the revised manuscript. However for closed-path system we also use the term dilution correction. About the "spectroscopic correction", we would like to keep it, because it is used also in other papers and gas analyser manuals.

13) On page 13 line 7 the authors seem to be implying that frequency response corrections are ambiguous and not based on physical laws. I would suggest that this implication is largely untrue. Frequency response losses are very real physical processes and the algorithms used to correct for these losses are well reasoned for the conditions under which they apply. Similarly, the WPL term is a simplification of the actual process; it is also well reasoned and applies for most conditions we are likely to encounter – but it is not the true correction.

Yes, we agree and we have removed that sentence.

14) On page 13 line 24 the authors suggest a bias will occur as a result of an inappropriate cospectral model being used in the frequency response correction. As always, it must be the responsibility of the researcher to make sure that the applied correction is appropriate for the experimental conditions encountered. It seems that an appropriate software should allow the user the ability to choose a cospectral model appropriate for experimental conditions.

Ok, but that is in fact the message we want to give. Of course, it is the responsibility of the researcher to make a proper correction, but sometimes the researcher processing EC data is not a micrometeorologist, and he/she may be not aware of critical factors causing systematic biases, like in this case. We think that this sentence will be useful for the reader in the present form.

15) In the conclusions (lines 20- 22) the authors suggest "that a consistent choice of implemented methods for the post-field processing steps can minimize the systematic flux uncertainty due to the usage of difference software packages". I find this to be a very dangerous conclusion to draw. This conclusion implies that there should be an uniformity in all flux calculations. Such a pressure to conform is likely to suppress the creativity and the willingness of researchers to explore the inherent variability of experimental field research under the presumption that all situations are identical. While I agree that algorithms should be employed as their originators intended; the concept that a prescribed set of instructions must be applied for a researchers fluxes to be considered acceptable is wholly unscientific.

Here we disagree with the reviewer, because with this conclusion we are not suppressing the creativity of the scientists. Instead, it is just the opposite like it is mentioned in the next few lines: "Finally, it is recommended in the future to work towards more software inter-comparison studies, where new methods and corrections are validated across different type of compounds/instruments and ecosystems".

16) On page 15 line 19, Wilczak et al. 2001 did not present the sector-wise planar fit approach, only the planar-fit approach.

Ok, corrected.

17) Appendix A, 'Calculation of turbulent fluxes': You present three methods of determining turbulent fluxes (block averaging, linear detrending, and autoregressive filtering) as if there

were completely independent processes. However, for run based statistics block averaging is inherent in any of the approaches taken such that linear detrending and autoregressive filtering simply become filtering methods for removing additional low frequency energy.

This is true for block averaging and linear detrending method, but not for the autoregressive running mean filter. Block averaging and respective filtering is applicable to running mean filtering method only if the mean removal operation is applied along with running mean filtering. If the deviations from running means define directly turbulent fluxes and no mean removal is performed over averaging period, filtering due to block averaging does not apply. See for discussion e.g. Rannik (2001) and Massman (2001). We have rephrased it.

18) Appendix A, 'Time lag determination and adjustment': Only a general description of lag time determination is given. No details on how lag time is selected for the many many situations in which the cross correlation curve does not show the obvious lag correlation peak.

More details on the time lag were added.

19) Appendix A, 'Density correction': The authors indicate that this correction only applies for open path trace gas sensors. Actually, it applies for any sensor for which the sample gas is not at constant temperature, pressure or composition of other trace gas species. This should also be reflected in the processing chain shown in figure 2.

This is not true, since the correction is described in the Appendix also for closed-path gas analyser. We will modify the Figure 2.

20) An important aspect that is missing from the analysis in this paper is the comparison of sensible heat flux computation. It is a vital component to many of the term and correction applied to the trace gas calculations shown in this paper and as such should warrant similar analysis.

We agree with the referee that accurate sensible heat flux (H) calculations are needed, since H is an important term used in some corrections. However, since the paper is focusing on gas fluxes, we decided not to make it longer by adding detailed H comparison, but rather we added a few sentences about it to the Discussion in the revised manuscript. Indeed, there seems to be a small difference in H after spectral corrections especially at Siikaneva site (see Fig. 2 below), which most likely stems from the fact that in EddyUH a site specific cospectral model was used, whereas EddyPro uses the cospectral model from Moncrieff et al.(1997).

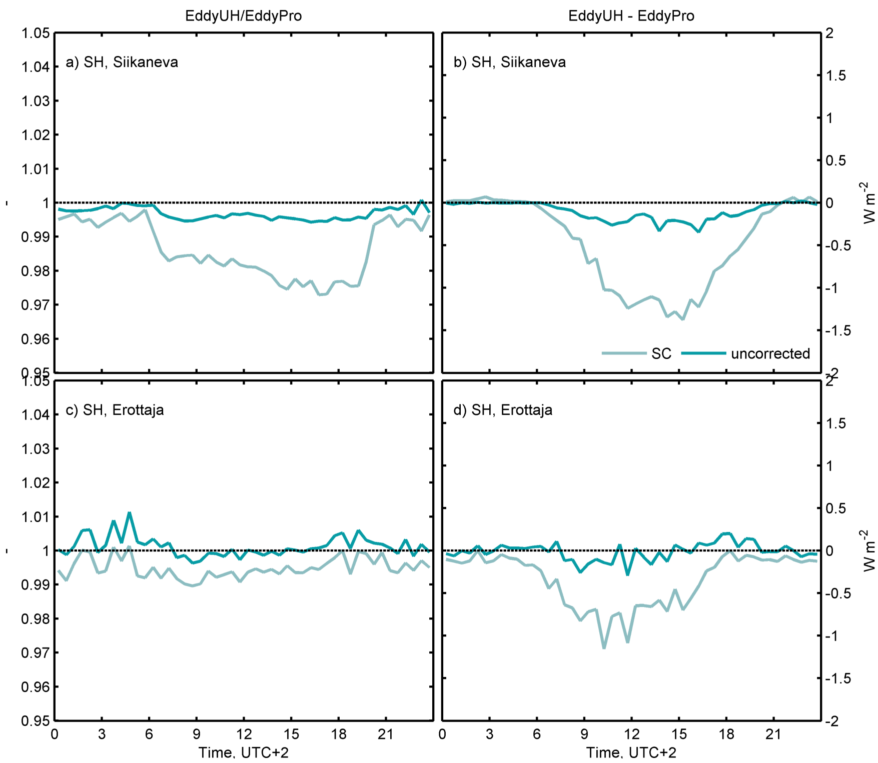

Figure 2. Comparison between sensible heat fluxes calculated by the two data processing programs. 'uncorrected' means sensible heat fluxes before spectral and SND-correction (van Dijk et al., 2004) and 'SC' show sensible heat flux comparison after spectral and SND-corrections, i.e. fully corrected fluxes.

21) There are several problems with the graphic in figure 1. A) the 'high frequency transfer function estimator' process has no outputs. B) the planar fit and footprint processes have no inputs C) the WPL and band broadening corrections have no representation D) Iteration is not presented.

We have modified the Fig 1 (which is now the fig. A1), to take into account the points A) and B). However the steps mentioned in C) and D) are related to Fig.2 (fig.1 in the revised paper).

References

Felber, R., Münger, A., Neftel, A., and Ammann, C.: Eddy covariance methane flux measurements over a grazed pasture: effect of cows as moving point sources, Biogeosciences, 12, 3925-3940, 2015.

Rannik, Ü., 2001. A comment on the paper by W.J. Massman `A simple method for estimating frequency response corrections for eddy covariance systems'. Agric. Forest Meteorol. 107, 241-245.

W.J. Massman, 2001. Reply to comment by Rannik on "A simple method for estimating frequency response corrections for eddy covariance systems". Agric. Forest Meteorol. 107, 247–251