

## ***Interactive comment on “Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3” by G. Fratini et al.***

**Anonymous Referee #2**

Received and published: 24 March 2014

The manuscript describes the evolution of the agreement between two widespread eddy-covariance software packages during an intercomparison that included stepwise improvements to the source codes and options of both. It is relevant (because an increasing number of synthesis studies involves flux results from different sites, groups and software packages, which would benefit from a processing that ensures comparable results), innovative (because the intercomparison is accompanied by an identification, classification and correction of the sources for differences found), and appropriate to AMT (because the processing from raw data to fluxes is an inherent component of the eddy-covariance measurement technique). My main, and only major, issue is that some of the conclusions drawn (not necessarily in that section but across the whole text) are not supported by the results, and occasionally even in conflict with them or with

C256

each other (see detailed comments). This may in part result from a somewhat unclear motivation and scope. An optional improvement related to this, from which the value of the paper as a citation reference could greatly benefit, might be to stress that users of either of the two software packages (ensuring a certain version number, processing scheme setting and dataset properties similar to one of the two tested datasets) may expect a closer agreement than between arbitrarily chosen software packages. Ideally, this would be supported by a probability density function or confidence interval of the flux differences after the last "round".

Detailed comments:

Abstract: The second part starting from "Our main aim" contains some summaries of misconceptions in the main text that are commented on later in this review - please screen carefully.

P2110L15-19: The difference between C1 and C4 requires some more explanation, or maybe it is better to bin them together. C4 also does not seem to be mentioned any more throughout the rest of the manuscript.

P2111L05: Since only C2 and C3 are given as alternatives, did you mean by C1 "C1 and C4" ?

P2112L05: I could not verify the claim that EddyPro is open source. Please note that the expectations of the average article reader (e.g. that open source software enables him to inspect and modify any part of the source code) should determine the classification of the software type; if you rely on a different definition of "open source" it would be more instructive to clearly mention if the source code is open or not, rather than use this term. In my previous (and maybe wrong) perception, ECO2S was and is open source, while EddyPro and TK3 are both not. If this is correct, state it clearly for all three software packages, or for none of them at all.

P2112L04-24: Please check this passage carefully for irrelevant advertising. For ex-

C257

ample, giving reasons why the use of EddyPro is rapidly increasing is speculation not withholding the criteria of scientific publication, unless there is a section about how user questionnaires on their reasons to use EddyPro have been evaluated.

P2108L15, P2113L17, P2118L19, P2118L25, P2119L23: The fact that the authors - and not the user - have full control of the source code, is mentioned at least five times across the manuscript - I wonder why? It is in conflict with the claim at P2112L05 (see comment above) of EddyPro being "open source". Also, it suggests that additionally having self-written software is of vital importance to anyone who aims at research-grade eddy-covariance processing and contributing to its further improvement. But this conclusion is not mentioned anywhere; instead in two places (see last comment) the wording even seems to suggest an inferiority of in-house software.

P2114L11: RMA is indeed one of the most reasonable options for this kind of comparison without a "true" reference. But since according to some definitions, RMA is rather an alternative to regression than a type of it (e.g. Webster, 1997, Eur. J. Soil Sci. 48: 557), this choice should also be mentioned in the figure caption to avoid misunderstandings (keeping the term "RMA regression" while doing so is fine with me). To avoid meaningless regression parameters caused by far outliers in the first round (Fig. 1), you might also want to check alternative methods from robust statistics. Or maybe use a significance test and don't show regression lines with an insignificant slope parameter, just the  $R^2$ . But on the other hand, as long as these parameters are only used to demonstrate that there are still a lot of far outliers at that stage (as currently done), I guess things can stay as they are. Note that one interesting conclusion is not yet mentioned: Since RMA slopes, unlike linear regression slopes, are not low-biased towards the chosen X variable but depend on the ration of the variances of both variables, the three low slopes in Fig. 1 indicate that in all these cases the variance of EddyPro results must have been several times larger than the one of the TK3 results.

P2115L23-P2116L02: Difficult to understand from this passage: Which software package supported which interpretation of the data initially, which one (if any) was then

C258

identified as erroneous by the authors, and which adaptations to which software package were then performed?

P2116L20: Maybe an example would be helpful here.

P2117L05: The term "possibly" raises some uncertainty as to whether and how iterations were performed (also does not seem to occur in table 2).

P2117L10: The decision to use the open-path dataset for ANOTHER three rounds rather than a repetition, though doubtlessly based on the experience of the authors and beneficial for the total agreement after the analysis, is a bit awkward and requires extra care in the interpretation and documentation. For example: Would the results of the first round on the open-path dataset have looked similar as the ones of the closed-path dataset, if the order of both datasets would have been reverted? Has it been ensured that the comparison results from the last round with the first dataset could still be reproduced (or at least didn't deteriorate) after the last round with the second dataset?

P2120L01: In fact the study demonstrates that the name of the software used will be of little help without the exact version number. By the way, providing here or in the abstract the version numbers of EddyPro and TK3 from which the "last round" agreement can be expected, would be highly useful.

P2120L02: The role of "ad-hoc" intercomparisons is somewhat unclear. Do you refer to what would have happened if a researcher was willing or able to only perform round 1 of the comparison? I agree that it would have been a pity not to take the chance to improve both software packages, but nevertheless that researcher would have obtained the correct quantification of the agreement of the used version numbers of both softwares, rather than a misconception to warn against.

P2120L03: Any software (including e.g. any new version release of an otherwise highly proven software) not undergoing systematic quality assurance will be a risk, no matter

C259

whether it is in-house or not (see also comment on P2108L15 ff.). Similarly, at the end of the abstract it would be sufficient to demand that researchers rely on extensively validated software, no matter whether it is "established" or not.

---

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 2107, 2014.