Reply to comments of Reviewer # 1:

We want to thank the anonymous referee for carefully reading our manuscript and for their helpful comments. We have organized the reviewer comments in a manner such that RxCn represents the nth comment of referee x, and RxSn the nth specific comment by reviewer x. We hope this will provide a clear basis for discussion during the further reviewing process. We are addressing the raised comments in a point-by-point way below:

The authors present an evaluation of four different REA β-factor estimation approaches. They evaluate the four approaches for the H_2O flux at three different sites with very different vegetation cover (forest, meadow, gravel). The comparison of the different approaches and the different sites are the main advantage of the study. Specifically, the performance of the β_w approach is included (which is rarely reported in the literature) and for the proxy β approach, the use of an overall constant β value is compared to a half-hourly adjusted value. A main result of the study is that the use of a constant β value per site (or individual β_w values, which show a quasi-constant behavior) is superior to the use of half-hourly determined proxy β values. However, the evaluation of REA approaches is less comprehensive than declared in the objectives. Only one scalar (H_2O / latent heat) is used for the approach validation, and only one scalar (T) is used as proxy. Moreover, the manuscript suffers from a number of additional shortcomings that need some substantial improvements before publication. They are listed in the following comments. Important: The line numbering of the manuscript is erroneous (non-sequential) on most pages, which made the review somewhat cumbersome. I use the true text line numbers in the following comments (not the ones indicated in the manuscript)

We are sorry for the additional work generated by erroneous line numbers. Thank you for making the extra effort.

R1C1) Only the performance of the REA approaches for the H_2O flux is tested in the present study. This is done after an initial deadband optimization (using the reference EC dataset) for the same test scalar. This leads to a certain lack of independence in the method validation. Although the CO_2 flux and its correlation with the other scalar fluxes is introduced in Sections 3 and 4.1, the REA evaluations for the CO_2 flux are unfortunately not presented. Alternatively CO_2 could have served as second proxy scalar option beside the temperature T (at least for some sites) as indicated in Section 2.3.

The authors should more prominently (in abstract and objectives) declare that they are evaluating the REA approaches only for H_2O fluxes. In addition they need to discuss better, whether and why they assume that the results also apply to other scalars, despite a sometimes low scalar correlation as exhibited in Fig. 3.

The reason why only the results for the H_2O flux are presented was to limit the analysis to a reasonable scope. Additionally, we decided to not present the CO_2 flux results because, for the gravel site (Antarctica), there is basically no measurable CO_2 flux due to lack of biological activity, which makes the interpretation difficult. However, we agree that, for method validation, considering another flux than the one for which the deadband size was optimized is required. Following the referee's suggestion, we propose adding an appendix (Appendix A), in which we present the hourly binned RMSE evaluation, which was done for H_2O in Fig. 9, but for the CO_2 flux. Alternatively, the below figure and interpretation could be included and discussed in the main manuscript. We would like to leave this decision to the editor. Regarding the second part of the comment, we state that the changes will be reflected in abstract and introduction.

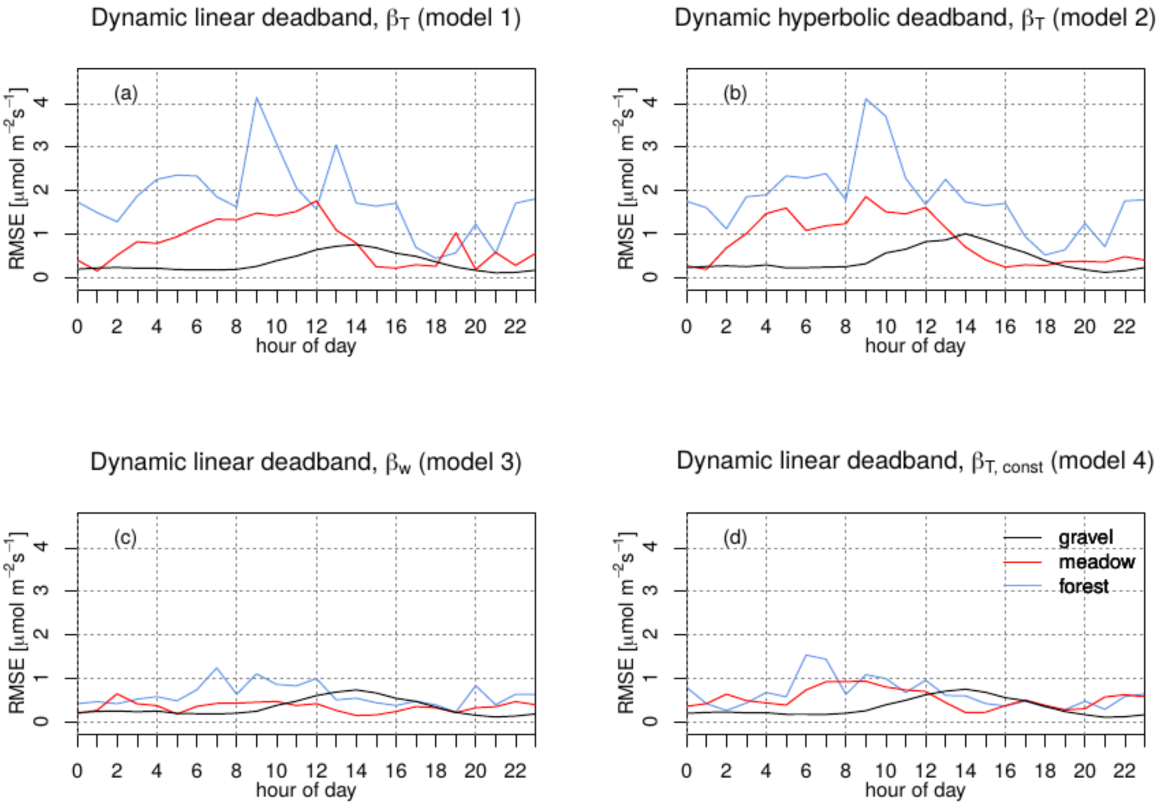

Fig. 11: Same as Fig. 9 but for the CO_2 flux. The gravel site results (solid black lines) should be regarded with caution as the magnitude of the CO_2 flux at this site is close to zero (compare to Fig. 2).

Interpretation: *The same findings that were drawn from the H_2O flux analysis are also apparent in the above figure: Both proxy approaches (panels (a) and (b)) result in higher values of the RMSE than the β_w (panel (c)) and the constant β (panel (d)) methods. The RMSE for both proxy approaches at the meadow site peaks during 13-14 UTC, the time when scalar-scalar correlation of sensible heat and CO_2 is lowest. At the forest site, the RMSE for the β_T approaches is highest when the magnitude of the CO_2 is largest. The RMSE for the gravel site is included in this figure even though the magnitude of the CO_2 flux is close to 0 throughout the daily course and thus no conclusions should be drawn from its RMSE.*

R1C2) I find it a bit misleading to use the index "0" for the β factor of the proxy scalar approach. Obviously (see scalar correlation analysis) it matters, which scalar is used as proxy. Therefore, it would be more informative and more consistent to use the scalar specific index "T" or "wT" for the proxy scalar approaches here.

We take this comment into account and agree with the reviewer using scalar specific indices are more clear. The adjustments have been made in the figures and the manuscript.

R1C3) The presentation of the β_w approach in Section 2.2 is a bit confusing in my view. It is not clear what the use of Eq. 5 is for a REA application. The factor "m" is a purely theoretical quantity that has no use for practical REA applications. Therefore the practical β_w approach evaluated in the present study should be clearly separated from theoretical considerations. In addition, the alternating use of " β" and " β_w " in this section is confusing. E.g. it is argued (P5, L5) that "the c'-w' correlation also affects β_w ". But this is contradicting the definition of β_w (Eq. 4) purely depending on the w-distribution.

We strive for our study to be helpful and understandable to users. Therefore, we thank the reviewer for pointing us to these inconsistencies in Section 2.2.

However, regarding the first part of the comment, we would like to argue that one needs to do both: A theoretical introduction, and an evaluation of the practical method. The latter cannot be done without the first, because the practical method needs to rest on a firm theoretical foundation. Our aim is to provide a brief yet comprehensive derivation, starting from theoretical considerations, in our Methods section.

We propose to rephrase the first part of Section 2.2 as follows:

“An alternative REA method was originally derived by Baker et al. (1992), and Baker (2000) provided a comprehensive derivation. It primarily rests upon the standard deviation of the vertical wind σ_w, and assumes velocity-scalar correlation. In brief, the flux is defined as:

$$\overline{w'c'} = m \cdot \sigma_w^2$$

where m is the regression-estimated slope of the w’ vs. c’ correlation. m can be approximated, using conditional sampling techniques, as:

$$m = \frac{\Delta\bar{c}}{\Delta\overline{w}}$$

This makes:

$$F_{REA} = \frac{\Delta\bar{c}}{\Delta\overline{w}} \cdot \sigma_w^2$$

and as a result, a β_w factor can be derived as follows:

$$\beta_w = \frac{\sigma_w}{\Delta\overline{w}}.$$

The scalar flux becomes directly proportional to the vertical wind speed's variance σ_w^2, and thus to the turbulence statistics. This approach combines elements of the flux-gradient and flux-variance similarity theories.

The requirements for this parameterization are (i) a linear relationship between c’ and w’ through the origin, as well as (ii) the Gaussian distribution of the vertical wind velocity fluctuations. If both are fulfilled, β_w = 0.63, however, usually, smaller values of the β_w parameter are measured (Katul et al., 2018). ”

R1C4) How can it be that the zero deadband calculations result in RMSE of about 20 mmol m-2 s-1 for the forest site in Figs. 5 and 6, when the fluxes themselves are only between 0 and 4 mmol m-2 s-1 (Fig. 2) and the flux ratios in the left panels are close to 1? This seems very unplausible and needs a detailed explanation.

Thanks for spotting this. The large RMSE compared to the median F_{REA}/F_{EC} ratio close to 1 was actually due to one single outlier. We decided to take the physical plausibility thresholds, which were applied to the EC fluxes, and also apply them to all simulated REA fluxes. This removes the outlier in question, and reduces the RMSE values for the forest site in Figs 5 and 6. However, the thresholding does not alter any of the other presented results significantly. The main finding presented in this section, i.e. that the proxy-based approaches result in a larger error compared to the β_w and $\beta_{T,const}$ approaches, remains still valid.

We propose to include the following explanation in Section 3.2, stating that the physical plausibility thresholds were applied to the simulated REA fluxes as well:

"In the final step, the same thresholds for physical plausibility which were applied to the computed EC fluxes were also used to remove unplausible REA flux estimates from the data sets. These thresholds were chosen individually for each scalar and each data set due to the wide range of meteorological and biochemical conditions covered in this study."

Updated Figs. 5 and 6:

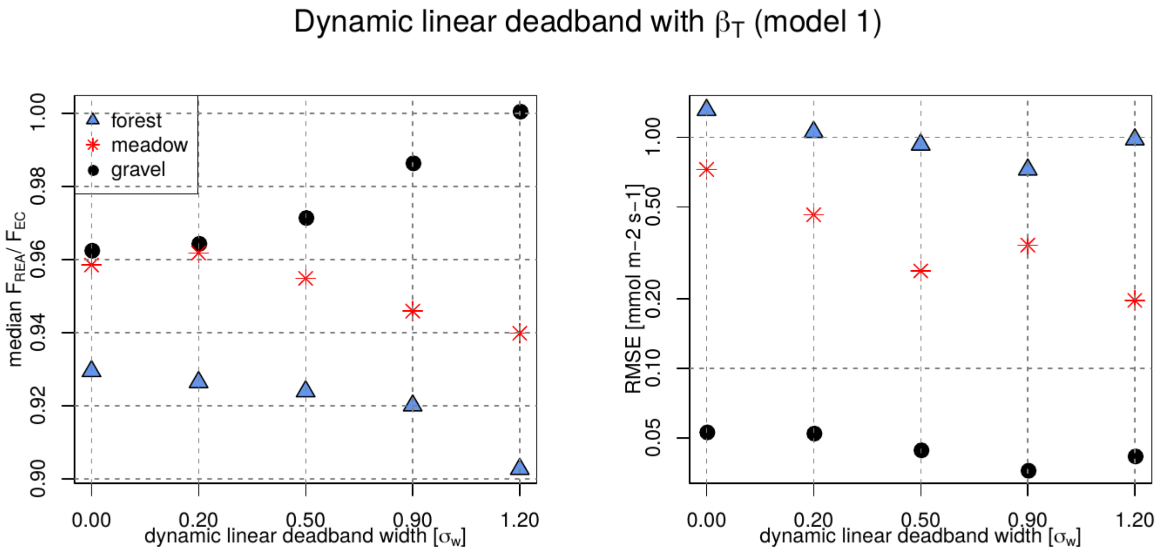

Fig. 5: Errors as a function of dynamic linear deadband width. The x axis is the scaling factor multiplied with the vertical wind standard deviation σ_w to define the deadband threshold. Left panel: Median F_{REA}/F_{EC} (latent heat flux simulated with sensible heat as a proxy) ratio for each of the simulated dynamic deadband widths; right panel: RMSE for each of the simulated dynamic deadband widths

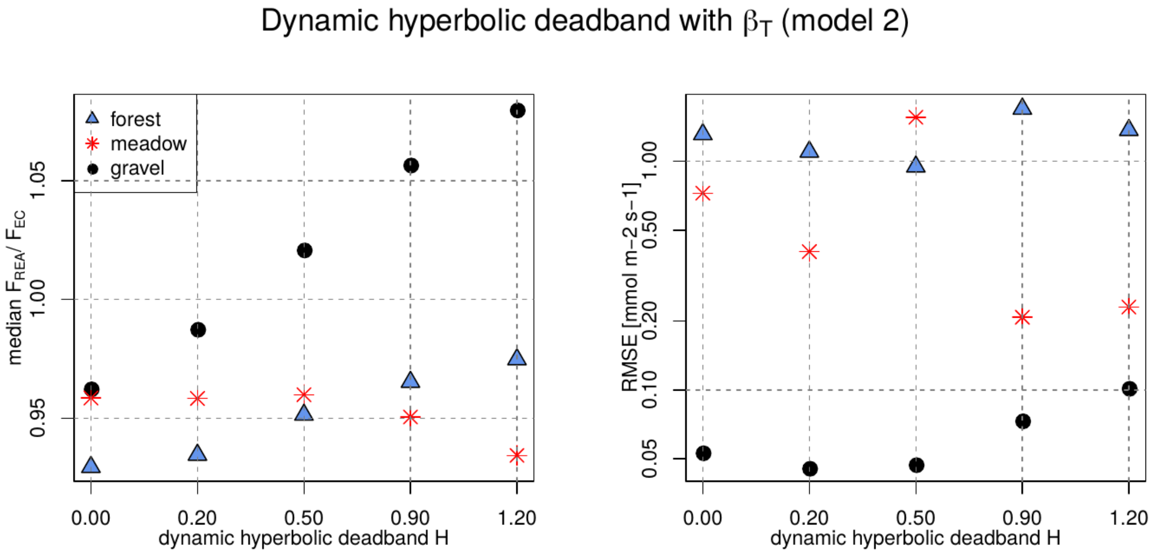

Fig. 6: Errors as a function of dynamic hyperbolic deadband size. The x axis is the H parameter in Eq. 10, which defines the deadband size. Left panel: Median F_{REA}/F_{EC} (latent heat flux simulated with sensible heat as a proxy) ratio for each of the simulated dynamic deadband sizes; right panel: RMSE for each of the simulated dynamic deadband sizes

R1C5) The resulting $\beta_{0,const}$ values and the average β_w values for the three different sites should be listed in a Table, so that other researchers can compare them to their own results.

We thank the reviewer for this good suggestion. As recommended, we are adding a table, in which the β values found for each of the sites and methods are listed for the respective optimum deadband sizes:

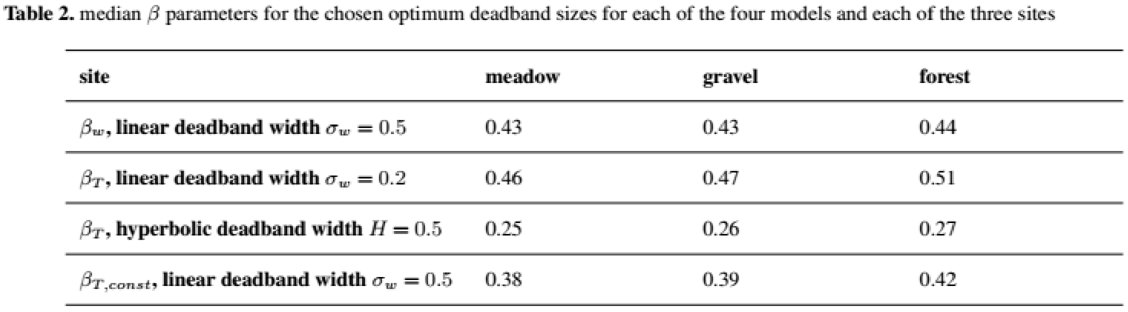

Table 2. median β parameters for the chosen optimum deadband sizes for each of the four models and each of the three sites

site	meadow	gravel	forest
β_w, **linear deadband width** $\sigma_w = 0.5$	0.43	0.43	0.44
β_T, **linear deadband width** $\sigma_w = 0.2$	0.46	0.47	0.51
β_T, **hyperbolic deadband width** $H = 0.5$	0.25	0.26	0.27
$\beta_{T,const}$, **linear deadband width** $\sigma_w = 0.5$	0.38	0.39	0.42

We propose to also add the following sentence to the end of Section 4.2:

„Table 2 summarizes the chosen optimum deadband widths for each of the four methods and gives the medians of the respective β parameters for each of the three sites. “

R1C6) For Figure 10 and 11 it is not indicated, which data are displayed. Are these all (valid) data for all three sites or only data from one site? This needs to be clearly stated in the Figure caption.

Thanks for bringing up this issue. In Figs 10 and 11, all valid data from all three sites are combined. The observations from all three ecosystems fall along the same lines, which suggests that e.g. the findings of β_w vs. kurtosis as a function of deadband size presented in the left panel of Figure 11 are ubiquitous.

For clarification, we are adding the following sentence to the caption of Fig. 10:

"This figure combines valid data points from all three sites."

and we are adding

"Valid data points from all three sites are combined in this panel.",

to the caption of Fig. 11.

R1C7) I have some problems when comparing the β_w results displayed in Fig. 10 (right panel) and Fig. 11. The zero deadband results in Fig. 11 show a considerable variation with the kurtosis and that the kurtosis systematically depends on stability. In contrast the β_w results in Fig. 10 show practically no variation, neither with stability nor within the bins.

We agree, at first sight these observations appear to contradict each other. However, when carefully evaluating these findings, it is an effect of the binning. Fig. 11 center panel only displays the bin medians, while Fig.11 left panel shows the unbinned 30-min data. Comparing the ranges of the w-kurtosis in these panels, one learns that the range between 3 and 4 (center) is much smaller compared to 2 and 5 (left). Within the approximate bounds of 3 and 4 (where most of the data are for all three sites), β_w for zero deadband also has a much smaller systematic variability. In combination with Fig. 10 right panel it means is that the bin median value of the w-kurtosis artificially exaggerates the stability dependence, since the within-bin variability is very large, leading to its effect disappearing in the effective β_w (Fig. 10, right) and F_{REA}/F_{EC} (Fig. 11, right) findings.

We have added arrows for the IQR of kurtosis and z/L to the center panel of Fig. 11 to make this point more clear, and updated the figure description:

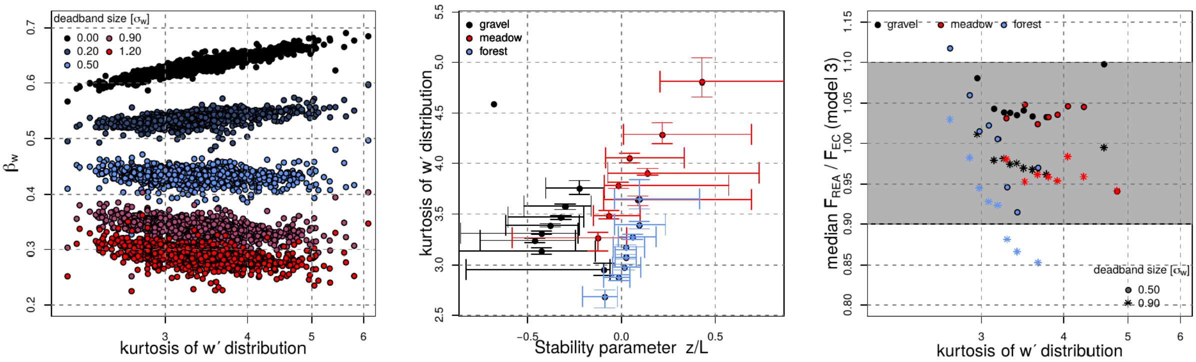

Fig. 11: Left panel: βw as a function of w' kurtosis for different deadband widths (not binned). Valid data points from all three sites are combined in this panel. Center panel: the stability parameter z/L as a function of the w' kurtosis. Data were binned into eight kurtosis bins with equivalent number of data points. Only bin medians are displayed, arrows mark the IQR. Right panel: Median F_{REA}/F_{EC} as a function of w' kurtosis for the optimal deadband widths, 0.9 σ_w and 0.5 σ_w, which were determined by Baker (2000) and in this study. Data were grouped into the same kurtosis bins as in the center panel.

We propose to add the above explanation to the text and restructure the combined discussion of Figs. 10 and 11 as follows to make it more logical:

"It was pointed out in previous REA studies that β_w scales with the fourth central statistical moment of the vertical velocity perturbations' distribution by altering the w' vs. c' relationship. We therefore investigated the impact of the w' kurtosis on the β_w factor for different linear deadband sizes.
Katul et al. (2018) found that two different factors, which both depend on z/L, contribute to β_w and whose impacts can cancel out if their magnitudes are similar. The first effect, leading to an decrease of β_w with increasing (positive) z/L, depends on the excess kurtosis, or flatness factor of the w' distribution. The second effect, resulting in an increase of β_w with increasing z/L, is a result of the transport efficiency e_T (Wyngaard and Moeng, 1992), as well as source strength and asymmetry in the w' distribution. The superimposition of these two processes

could be an explanation why there is no clear dependence of β_w *on dynamic stability visible in Figure 10.*

The relationship between the w' distribution's kurtosis and the β_w *factor is illustrated in Figure 11: consistent with Katul et al. (1996, 2018), the* β_w *factor without deadband increases as a function of w' kurtosis (Fig. 11, left panel). The plot collapses data from all three ecosystems onto a single linear relationship. This finding suggests that the turbulence statistics are ubiquitous despite the significant differences in climate and surface characteristics across the three ecosystems. The increasing linear trend becomes less pronounced when deadbands are applied.*

Kurtosis is in turn expected to be related to dynamic stability, when changes in turbulence statistics and diabatic conditions lead to non-Gaussian distribution of w'. As a result, the kurtosis of the w' distribution becomes different from 3, which is the value for a Gaussian distribution. In the center panel in Fig. 11, w' kurtosis is plotted against the stability parameter z/L. The right panel of Fig. 11 displays the resulting median F_{REA}/ F_{EC} *as a function of w' kurtosis. Only the model results for REA applying a linear deadband with widths of 0.5 and 0.9* σ_w *are displayed here for improved visibility. While no clear trend is observed at the grassland site, and only a slightly negative trend is visible for the gravel site, we can detect a strong decrease of the median* F_{REA}/ F_{EC} *as a function of w' kurtosis for the forest site. However, as is indicated by the shaded area in the rightmost panel of Fig. 11, most points lie within the boundaries of +/- 10 %. Only the bins with the highest and lowest kurtosis classes at the forest site are outside of this range. These error bounds are of the magnitude as the error assumed in EC applications. We suspect that the large excursions from Gaussian statistics for the forest site are caused by coherent structures forcing cross-canopy vertical exchange, which are a dominant flow mode in the forest flows documented for this site (Thomas et al. 2007a, 2007b).*

At first sight, it is puzzling why the β_w *model without deadband (deadband size 0.00) in Fig. 11 shows a considerable variation with the kurtosis, which in turn is related to stability, but basically no dependence of the* β_w *factor on stability can be seen in Fig. 10. This effect is due to the binning: The values in the center panel of Fig. 10 are bin medians of the kurtosis, while in the left panel, the unbinned 30-minute data are shown. Comparing the ranges of the w-kurtosis in these panels, it becomes apparent that the range between 3 and 4 (Fig. 10 center) is much smaller compared to the range between 2 and 5 displayed in the left panel of Fig. 10. Within the approximate bounds of 3 and 4 (where most of the data are for all three sites),* β_w *for zero deadband also has a much smaller systematic variability. Combining these insights with Fig. 10, it means that the bin median value of the w-kurtosis exaggerates the stability dependence, since the within-bin variability is very large, leading to its effect disappearing in the effective* β_w *(Fig. 10, right) and* F_{REA}/ F_{EC} *(Fig 11, right) results. Our findings indicate that the variation of the* β_w *factor with the turbulence statistics seems to have no significant impact on the flux estimate.”*

We want to keep the rather theoretical discussion of w-kurtosis to provide an observation for a potential explanation as to why beta values vary. Again, we believe it is important to understand the theoretical foundation of approaches, irrespective of their impact on the practically applied method.