Reply to comments of Reviewer # 2:

We want to thank the reviewer for their careful reading and helpful comments, which clearly helped improve our manuscript. 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:

This paper presents a contribution to the evaluation of the β parameter required for the relaxed eddy-accumulation (REA) technique. This technique is used to measure land-atmosphere exchange of scalars for which analyzers fast enough to implement the eddy-correlation technique are not available. The study is based on fast observations of temperature, vertical wind and humidity, on three contrasting ecosystems during a few weeks. The authors have simulated a relaxed eddy-accumulator on the recorded time series, and have compared the resulting moisture flux estimated for different β models with the eddy-correlation value, the latter being considered as the true value of the flux.

R2C1) I am not convinced that BG is a suitable journal for such a study. The paper is technical, and does not offer any process analysis. In my opinion, AMT would be more appropriate. But I leave to the Editor(s) the settlement of this question.

We appreciate this comment. Our submission to BG was motivated by the journal's focus on the **interaction** between biological, chemical, and physical processes, which is basically the main concern of the flux measurement community. We were hoping to find a good platform to reach the audience interested the most in our results by publishing in BG. We leave the decision to the editor.

R2C2) There is an abundant literature on REA, β determination and sensitivity to various parameters. By the way the authors mention numerous previous studies in their paper. However, they do not clearly indicate what is really innovative in their study, what is a progress with respect to previous estimations/ models, etc.. For example, the detection limit and sensitivity of the analyzers is often an obstacle for trace species flux estimates.

The authors indicate in their abstract that "For conditions close to the instrument detection limit, the β_0 models using a hyperbolic deadband are the optimum choice.", but this statement is not really supported by a study in which time series would have been degraded to simulate a less performing analyzer.

To date, the existing papers investigating the REA method either focus on only one or a smaller selection of β approaches, and/or a limited selection of sites (mostly only one site). These limitations make transferring those results to an arbitrary (new) site difficult, which aggravates the choices a REA user needs to make. To our best knowledge, no study has compared across these different methods. Hence, we included all β approaches across a very broad range of contrasting sites, this is the main innovation and contribution. We are proposing to add the following sentence to the Abstract to make this point clearer:

"To our best knowledge, this is the first study inter-comparing these different approaches over a range of different sites."

Furthermore, at least in the atmospheric chemistry community, the β_w method is not well known and our contribution demonstrates that it is capable of yielding results as good as (or actually even superior to) the better-known proxy approaches. Another interesting outcome of this study is that the use of the constant β_T factor performed better than β_T factors which are adjusted for each sampling period.

We are rephrasing part of the Abstract as follows:

"With respect to overall REA performance, we found that the β_w and constant $\beta_{T,\,const}$ performed more robustly than the proxy-dependent approaches."

Regarding the second part of the comment, we cannot possibly simulate the variety of analyzers all subject to different detection limits in our analysis. The latter a potential user is most familiar with, but she/he may require guidance on the β model. The ultimate choice which uncertainty (β approach, or analyzer detection limit) weights more heavily, is up to the user.

R2C3) The authors evaluate several models for the parameter β. It is sometimes difficult while reading the paper to clearly understand to what model it is referred to. For example, it is written in the abstract "We tested a total of three different REA models for the β factor...", whereas in the text 4 models are analyzed. In section 2.5, when the 4 models are presented, the corresponding relevant equations should be recalled. Furthermore, since they are numbered (#1... #4), the reference to the corresponding number should be systematically given both in the text and the figures.

The authors agree that the wording in the abstract is misleading. We rephrased the abstract text as follows to make it more consistent:

"We tested a total of four different REA models for the β factor: The first two methods, referred to as model 1 and model 2, derive β_T based on a proxy for which high-frequency observations are available (sensible heat T_s). In the first case, a linear deadband is applied, while in the second case, we are using a hyperbolic deadband. The third method, model 3, employs the approach first published by Baker et al. (1992), which computes β_w solely based upon the vertical wind statistics. The fourth method, model 4, uses a constant $\beta_{T,\,const}$ derived from long-term averaging of the proxy-based β_T factor. Each β model was optimized with respect to deadband type and size before intercomparison."

Furthermore, we want to thank the reviewer for the idea to recall the relevant equations in section 2.5, which will certainly help the comprehensibility of our methods description. We are adding them where we list the different model setups:

"

- *Model 1: β_T (Eq. 2) using the sensible heat as proxy and dynamically adjusted linear deadband scaled with σ_w (Eq. 9)*
- *Model 2: β_T (Eq. 2) using the sensible heat as proxy and dynamically adjusted hyperbolic deadband scaled with σ_w (Eq. 10)*
- *Model 3: β_w (Eq. 7) using a dynamically adjusted linear deadband scaled with σ_w (Eq. 9)*

- *Model 4: $\beta_{T, const}$ (Eq. 8; median over the complete field experiments) using the sensible heat as proxy and dynamically adjusted linear deadband scaled with σ_w (Eq. 9)"*

R2C4) The paper is confusing in several parts regarding the use of density vs. mixing ratio to express concentration. This is an important question, since we know from 40 years that density fluctuations have a considerable impact on flux estimates. This question is as crucial for REA as for eddy-correlation fluxes. See also my specific comments relative to this question below.

We want to thank the reviewer for raising this important point. We are addressing this issue in more detail below (see our answer to comment **R2S3**).

R2C5) The authors present CO_2 fluxes in their set of observations, but they do not use them to evaluate the β models. Only water vapour fluxes are analyzed. Why?

This issue was raised by the anonymous referee #1 as well, and was addressed in the response to **R1C1**. In short, CO_2 was excluded to keep the scope small, and because there is basically no detectable CO_2 flux at the Antarctic site. We are presenting a short evaluation of simulated CO_2 fluxes in the response to **R1C1** and propose to add it as an appendix, or include it into the Results section.

Here is the response to **R1C1**:

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.

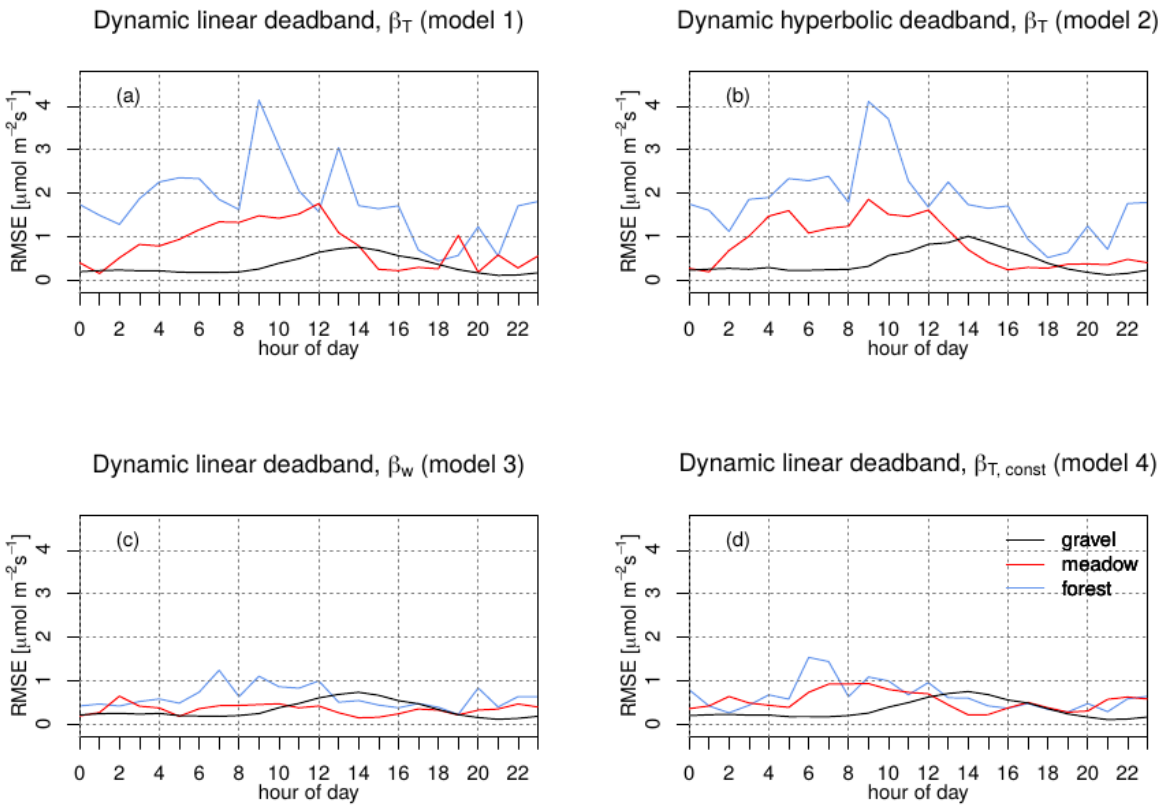


Fig. S1: Same as Fig. 9 but for the CO_2 flux. The gravel site results (solid black lines) should be regarded with caution as the magnitude of the CO_2 flux at this site is close to zero (compare to Fig. 2).

Interpretation: *The same findings that were concluded 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.*

Specific comments and drafting matter:

R2S1) P. 1, line 22: "To explain these surprising differences,...". To what differences is it referred to?

Thanks for spotting this, we understand that our text reads a little incoherently here. We refer to the difference between the β_w and β_T approaches with respect to their dependence on atmospheric stability. We propose to rephrase this part of the abstract as follows:

"To explain why the β_w method seems to be insensitive towards changes in atmospheric stability… "

R2S2) P. 2, lines 15-16: EA or REA techniques are NOT adapted for highly-reactive species, because concentrations might evolve under chemical reactions occurring during the accumulation period of time. For such species, disjunct eddy covariance technique can offer an interesting alternative.

We agree, however, DEC only offers an advantage for highly reactive species if the residence time in the system is small. While we think that adding a DEC simulation would be outside the scope of this paper, we agree that this method should be acknowledged in the introduction in order to give a more comprehensive overview. We are rephrasing the relevant part of the introduction as follows:

"However, such sensors are not available for all trace gases of interest, particularly for reactive species with brief atmospheric lifetimes. In these cases, Disjunct Eddy Covariance (DEC, Rinne and Ammann, 2012), i.e. non-continuous sub-sampling of the concentration and wind data series, offers one alternative to overcome this problem. Eddy Accumulation (EA) methods provide another solution for estimating the net flux of chemically more stable atmospheric species existing at very low concentrations. This technique was originally proposed by Desjardins (1972, 1977): In EA, a system of fast switching valves collects air into two separate reservoirs..."

R2S3) P. 2, line 26: The term "concentration" is ambiguous. It must be clearly indicated whether it means "density" (expressed in e.g. kg/m^3, or mol/m^3) or "mixing ratio" (expressed in e.g. kg/kg or mol/mol). Note that the mixing ratio value is conserved when temperature, pressure or density vary, which is not the case for density. When using an REA system, the mixing ratios have to be measured in the reservoirs at the end of the accumulation period because densities can have been modified under the variation of temperature and pressure conditions in the reservoirs. Similarly, if c is expressed as a mixing ratio, the correct form for equation (1) should involve the mean air density (see e.g. Bowling et al. 1999).

We agree, the entire volume expansion/ contraction argument leads to a significant correction term incorporated into the WPL correction (see Detto, M., & Katul, G. G. (2007). Simplified expressions for adjusting higher-order turbulent statistics obtained from open path gas analyzers. Boundary-Layer Meteorol., 122(1), 205–216.). Please also see the reply to **R2S21**, where we refer in more detail to the density correction applied.

We here use densities throughout the manuscript, these observations were collected with open-path analyzers. Since we focus on investigating the theoretical aspects of the REA approaches, we leave the physically correct conversion of densities into mixing ratios to the informed user/reader. Hence, our results are not affected by this important distinction as we stay with densities throughout.

We are adding the following sentence to the part of the Introduction where we introduce the REA method:

"Note that the term "concentration"' refers to densities (expressed in e.g. mmol m^{-3}) throughout this paper."

R2S4) P. 4, equation (4): Δw is not defined.

Thanks for spotting this! We added the following explanation:

"$\Delta\overline{w}$ is the difference of the mean vertical wind while sampling into the up- and downdraft reservoirs, respectively."

R2S5) Section 2.4: A discussion about the symmetrical vs. asymmetrical deadbands is missing.

Asymmetrical deadbands are only relevant in case of non-Gaussian flow and concentration statistics, which can be gleaned from investigating skewness and kurtosis as the 3rd and 4th central statistical moments. We here include an in-depth discussion of the kurtosis.

We are considering this comment and are addressing this in point in Section 2.4 as follows:

"Here, we only consider symmetrical deadbands, presuming symmetrically distributed flow and concentration statistics. Effects of non-Gaussian distributed w' and s' can be gleaned from investigating higher central statistical moments."

R2S6) P. 5, line 19: "applying a linear deadband to w"': please explain what is meant by "linear".

You are right, the term 'linear deadband' has not been introduced thoroughly in our text. Thank you for pointing us to this shortcoming. With a linear deadband, we mean a deadband linearly scaling with σ_w (fixed fraction 'a', linear equation: deadband size = a*σ_w + 0, where b=0 is the intercept).

We are including a more in-depth explanation:

"When applying a linear deadband to w' (left panel in Fig. 1), no sample is taken if the magnitude of w' is below a certain threshold. This threshold can be held constant or adjusted dynamically in time. Dynamical adjustments are often done by scaling with the standard deviation of the vertical wind σ_w. The linear deadband appears as two horizontal lines in the quadrant plot in Fig. 1 (left panel), defined by the linear equation

$$a \cdot \sigma_w + 0$$

where a is a constant."

R2S7) P. 5, line 22: "the deadband being proportional to the integral strength of the turbulent diffusive process". This is unclear. What is the "integral strength"?

The vertical velocity variance mathematically is the integral of the w-power spectrum. This nomenclature is common in micrometeorology (e.g. compare integral turbulence characteristics: σ_w/ u*, or integral turbulence intensity: σ_w/U).

R2S8) P. 5, lines 26-27: “Hyperbolic deadbands aim to exclude eddies with little flux contribution and maximize the concentration difference between the two sampling reservoirs.”. This is not specific to hyperbolic deadbands since the same can be said regarding constant deadbands.

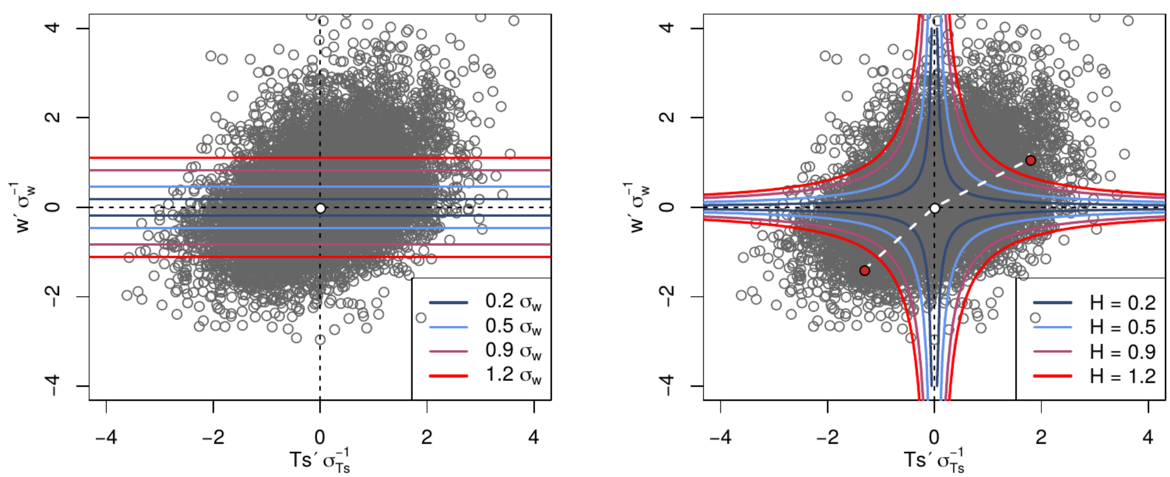
Agreed, linear deadbands have the same goal, however, for hyperbolic deadbands the above statement is even more applicable: There is a distinct mathematical difference. Hyperbolic deadbands filter in the w'c' plane (i.e. in the plane of instantaneous flux contribution given by the instantaneous cross-product), while constant deadbands filter only in the w' (or σ_w) space. In case w' is large and c' very small, then the flux contribution is small, and the sample may be discarded by the hyperbolic, but considered by the linear approach.

We suggest to rephrase the sentence as follows:

“Hyperbolic deadbands are specifically designed to exclude eddies with little flux contribution and maximize the concentration difference between the two sampling reservoirs.”

R2S9) Figure 1: I suggest to plot the symbols with a single colour, and add lines of different colours to indicate the thresholds, rather than plotting couloured symbols. As it stands the Figure is ambiguous: when we look at, for example, the light blue 0.5 σ_w threshold, all the red and brown dots should also be included in this class.

Thanks for the hint. We changed the figure accordingly:



The adjusted Fig. 1: Schematic quadrant plots to visualize the application of linear (left) and hyperbolic (right) deadbands. Different colors show which data points are included for different deadband sizes. The white dot marks the origin in both panels. In the right-hand panel, solid red dots mark the mean w'/σ_w and mean $T_s'/\ \sigma_T$ for up- and downdrafts when a hyperbolic deadband with H = 1.2 is applied. The white dashed lines in the right-hand panel connect the red dots with the coordinate system origin. The deviation from 180° of the angle spanned between these lines is a measure for the asymmetry of the sample distribution.

R2S10) Figure 1: What represent the black lines?

The dashed black lines represent the coordinate system axes through the origin; The continuous black line (not present in the updated version) represented the line through the two dots [mean(Ts’ (w’ < 0));

mean(w'(w' < 0))] and [mean(Ts' (w'>0)); mean(w'(w' > 0))], respectively, computed over all samples (no deadband applied). We decided to remove it from the updated version of Figure 1 because it did not add much value to this conceptual plot.

R2S11) Equation (1): Defined in that way, H is simply the correlation coefficient between w and s. The correct expression is to remove the overbar in (1) and mention that only the samples for which abs(H) is higher than a given threshold are retained.

You are right, we are changing the equation in question so that it matches Eq. (2) in Bowling et al. (1999):

$$H = \left| \left(\frac{w'}{\sigma_w} \right) \left(\frac{p'}{\sigma_p} \right) \right|$$

Written in this form, the relation to the hyperbolic graphs in the right panel of Fig. 1 also becomes more obvious. Please note that we have changed "*s*" to *"p*", to address the issue raised in **R2S12**.

R2S12) Equation (1) and throughout the text: "c" is used for an unspecified variable, but later on it represents the CO_2 concentration. This is confusing.

Thanks for the hint, which will certainly improve the comprehensibility of our manuscript. We decided to use "c" for CO_2, "s" for the scalar of interest, and "p" for the proxy scalar throughout the paper.

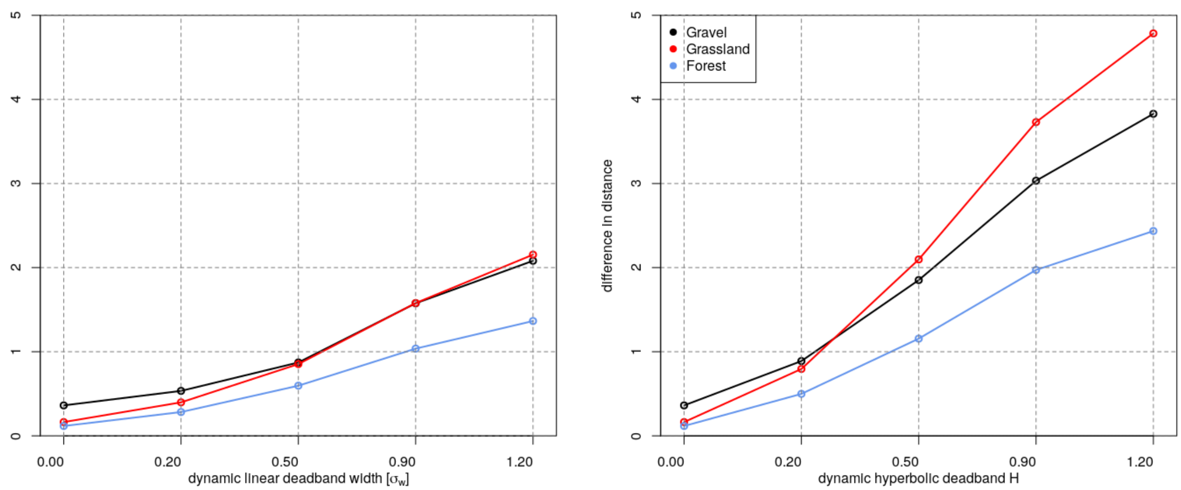
R2S13) Page 6, line 5: what is "n"?

n is the number of valid samples. We suggest to rephrase the part of the text in question as follows:

"The use of large deadbands must be done with caution because they exclude a significant fraction of the data from being sampled. As a result, the random sampling error, which is related to $1/\sqrt{n}$ *, can be increased due to the decreased sample size n."*

R2S14) Page 7, last line of 2.4: I am not convinced that the angle reflects the asymmetry of the sample distribution. A highly-skewed sample distribution would be represented by dots at a really different distance from the (0,0) coordinate origin, but these dots could be aligned with the origin.

We agree that the deviation from the straight (180°) line (which we are using in our analysis) is not sufficient to describe the asymmetry of the sample distribution. We had a deeper look into our data, and computed an additional asymmetry measure as suggested in the comment. Below figure displays the difference in distance to the origin; i.e. we compute the distances of the red dots to the coordinate origin (0,0; white dot) in Fig. 1, and then look at the difference of these distances, which represents skewness of the sample distribution.



Asymmetry measure (difference in distance of the two connecting lines to the origin) as a function of deadband size for linear deadbands (left) and hyperbolic deadbands (right).

A similar pattern as visible in the two bottom panels of Fig. 4 emerges: The asymmetry increases significantly with increasing deadband size. One difference to Fig. 4 is that the values for the gravel site lie in the same range as for the other two sites, which was not true for the asymmetry expressed by the deviation from 180°.

However, as this evaluation does not add any new insights into the asymmetry dependence on deadband/ sample size, we propose to not include this figure in the paper. Instead, we suggest to simply rephrase the sentence introducing the asymmetry measure:

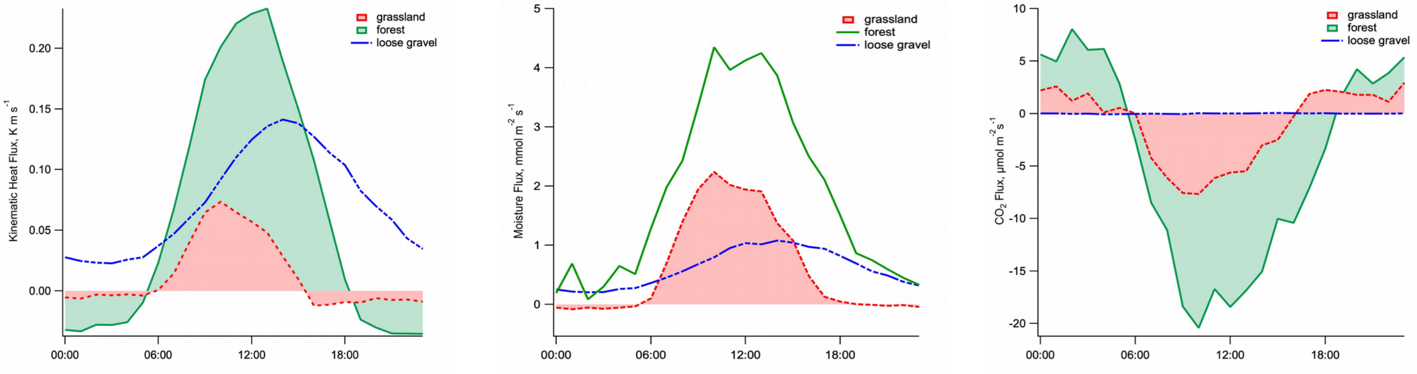
*"The asymmetry is shown as a white dashed line in the right panel of Fig. 1 containing a bend. This bend, which can be expressed as an angle deviating from 180°, is **one** measure for the asymmetry of the sample distribution."*

R2S15) Table 1: I suppose there is a mistake in the unit of roughness length.

Thank you for spotting this error! Indeed there was an error in our code to compute z_0 according to Eq. (2) in Panofsky (1984). The values are now 0.18 m, 0.06 m, and 4.87 m for the grassland, the loose gravel, and the forest site, respectively. We changed the numbers in Table 1 accordingly.

R2S16) Fig. 2: Replace "sensible heat flux" with "kinematic heat flux" and "latent heat flux" with "moisture flux"; or convert units into Wm^{-2}.

Figure 2 has been updated accordingly, thanks for the hint:



The updated Fig. 2

R2S17) Page 9: "During calm conditions, the region is dominated by a strong near-surface temperature inversion.". This is surprising, because Fig. 2 shows a positive (upward) heat flux, which is in conflict with a near-surface temperature inversion.

We are addressing this comment together with comment **R2S18** because they refer to the same issue.

R2S18) Page 9: "Strong katabatic winds draining the polar plateau frequently disrupt this inversion.". This is not consistent with the wind values given in Table 1.

We are addressing this comment together with the previous comment (**R2S17**). Both mentioned text passages refer to polar night conditions in Antarctica, which were not encountered during the field campaign used in our study. We removed the sentences from the manuscript.

We want to thank the reviewer for noticing this inconsistency. It is correct that during our campaign, which took place in Austral summer, we only encountered upward directed heat fluxes and no surface inversion at all.

R2S19) Page 10, line 13: Please explain what is a "dynamically determined lag.".

The sentence "Covariances were maximized by shifting the scalar time series relative to that of the vertical velocity by a dynamically determined lag.", refers to a step during data processing, where the time series of scalar concentration is shifted to achieve maximum cross-correlation with the vertical wind time series. This shift is determined dynamically, i.e. individually for each sampling period. After shifting the time series, the initial covariances can be calculated (see Foken, 2008, Micrometeorology, p. 109).

We are adding the following sentence to make this clearer in our text:

"This means that, for each sampling period, the scalar time series were shifted to achieve maximum cross-correlation with the vertical wind time series (Foken, 2008)."

R2S20) Page 10, line 20: Please explain what is meant by "additional hard thresholding was applied.".

We applied physical plausibility thresholds to filter the data for unphysical outliers. These thresholds were different for each scalar and each data set, due to different biochemical and meteorological conditions, and different measurement systems used. More specifically, the thresholds were defined as follows in our code (for the Dry Valleys/ gravel site "DRYVEXA", the forest site "WS2016", and the meadow site "ExpMM2015", respectively):

```
if (d=="DRYVEXA"){phys=c(0,400,0,300,-0.020,0.010)} # sensible heat, latent heat, co2

if (d=="WS2016"){phys=c(-100, 550,-100,400,-0.010,0.010)}

if (d=="ExpMM2015"){phys=c(-100, 200,-80,250,-0.020,0.010)}
```

To improve the readability of our text, we added an explanatory sentence to Section 3.2, which combines this issue with another comment raised by Referee #1 (**R1C4**):

"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."

R2S21) Page 10, lines 25-27 (also in line with comment#4 above): "To this end, molar densities were multiplied by the ratio of the instantaneous to mean density of moist air q <q>−1. EC fluxes were computed using the common post-hoc density correction (Webb et al., 1980).". This is unclear. Please check carefully. When one computes eddy-correlation fluxes with the scalar fluctuation expressed in mixing ratio unit (or any other proportional unit), the so-called WPL correction is not to be done. What I understand here is that the authors start with a mixing ratio, convert it to a quantity proportional to a density, and eventually apply the WPL correction (one step back, one step forward...).

Thank you for bringing up this point. It is always beneficial to review the data processing steps and check for consistency.

We start with densities (in mmol m^{-3}), which are being corrected by an *ad-hoc* density correction. *Ad-hoc* means that the correction (Detto and Katul, 2007) is directly applied to the high-frequency time series.

Starting from Eq. (3) in Detto and Katul (2007):

$$\rho'_{c,ext} = \frac{n_c}{n_a} \rho'_a$$

with

$$\frac{n_c}{n_a} = \frac{\overline{\rho_c}}{\overline{\rho_a}}$$

where overbars indicate means and primed quantities are fluctuations; ρ_c is the density of a scalar c, ρ_a the density of dry air; $\rho'_{c, ext}$ are the fluctuations of scalar c due to fluctuations in external conditions (mainly due to changes in air temperature and water vapor density). n_c is the number of molecules of scalar c.

Combining the above equations and solving for the "correct" density, $\rho_c - \rho'_{c, ext}$ (which does not contain the fluctuations of external conditions), leads to the correction mentioned in our text, e.g. for CO_2:

$$CO_{2\,corr} = \frac{\overline{\rho_a}}{\rho_a} \cdot CO_2$$

This correction leaves the density units untouched (the ratio of fluctuating to mean pressure is dimensionless).

This *ad-hoc* correction is an alternative to the *post-hoc* WPL correction. However, since we sample in the instantaneous w's' planes, we need to account for the contraction-expansion argument (Detto and Katul, 2007) for every single sample. Our EC fluxes (to compute the F_{REA}/F_{EC}) ratios in turn were corrected using the default *post-hoc* WPL correction.

In summary, the first part of the quoted text in the comment refers to the *ad-hoc* density correction, which is needed for the REA simulator. The WPL correction was only applied for computation of the EC fluxes. We agree that this passage reads somewhat confusing and have tried to rephrase it in a more understandable way as follows:

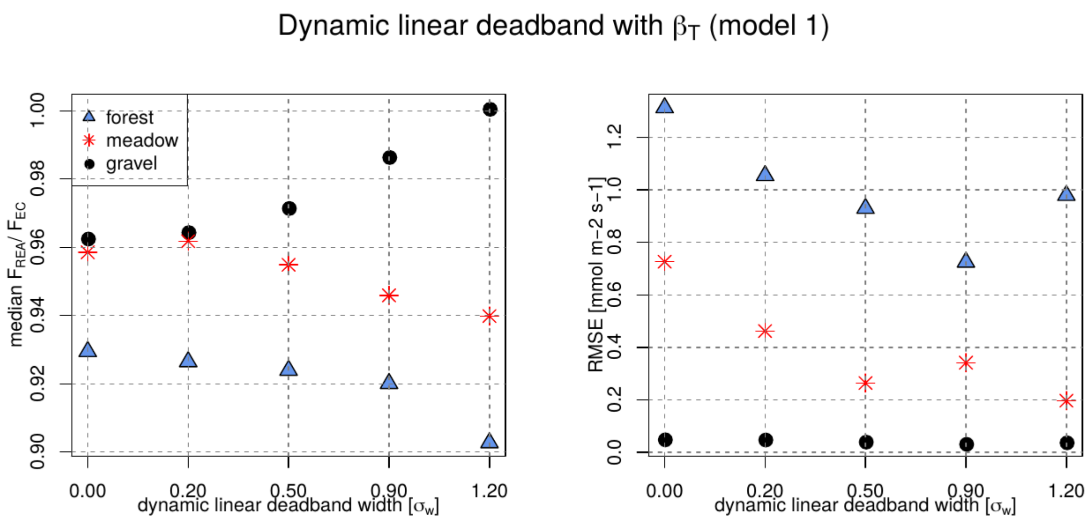
"Since simulating REA sampling requires selecting individual high frequency data from a continuous time series and computing density-corrected scalar higher-order moments, an ad-hoc density correction was applied to the water vapor and carbon dioxide molar densities (Detto and Katul, 2007) prior to flux computations. To this end, molar densities were multiplied by the ratio of the instantaneous to mean density of dry air $\rho_a \overline{\rho_a}^{-1}$ *. This correction removes the density fluctuations due to changes in external conditions. EC fluxes were computed using the common post-hoc density correction (Webb et al., 1980)."*

R2S22) Figure 4: "valid samples" is not the most appropriate terminology. Rather "selected samples", or something equivalent…

Agreed, we changed the figure accordingly.

R2S23) Figure 5: Why a logarithmic scale on the right panel ? RMSE unit is missing.

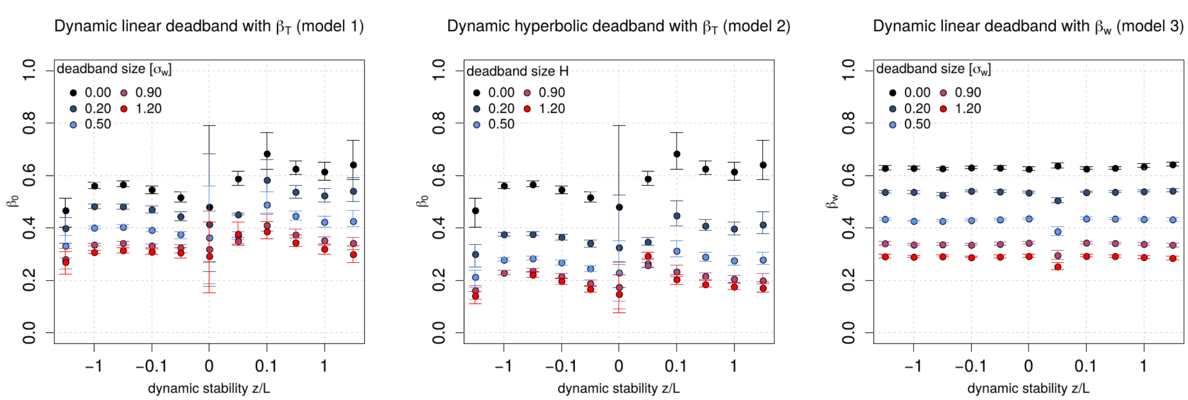
Due to the large variety of meteorological conditions covered by the three data sets, we obtain a large range of RMSE values. We felt that a logarithmic scale makes the plot clearer than a linear scale. However, after more thorough data cleaning in response to a comment by referee #1, the ranges covered by the RMSE are not that large anymore. We are changing the y axes of Figs. 5-8 to linear scale, see the updated Fig. 5 below. Thank you also for spotting the missing unit. We updated the figure accordingly:



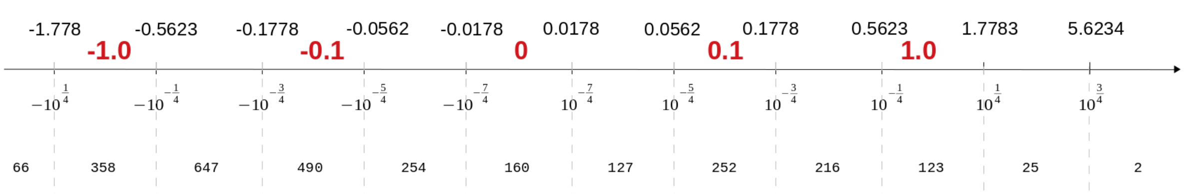
The updated Fig. 5. The missing unit of the RMSE was added and the y axis scale in the right panel was changed to linear.

R2S24) Figure 10: It is unclear how the stability bins are defined. It seems that the neutral class is very large, encompassing stable and unstable conditions until abs(z/L)~0.1. There is therefore an evidence of overlapping between the classes.

Originally, we wanted to define a larger "neutral" class (with intended overlapping), however we see that this does not really provide any additional insight. We reduced the neutral stability class to z/L values between -0.0177 to + 0.0177 (or +/- $10^{-1.75}$), and removed the vertical dashed lines from Fig. 10:



The bins are defined as logarithmically evenly spaced classes of dynamic stability. We split the range of encountered values of dynamic stability as shown in the below schematic. By defining the boundaries as $-10^{0.25}$, $-10^{-0.25}$, $-10^{-0.75}$, … the centers of the bins (on a logarithmic axis) are the red values, which correspond to the labels on the x-axis of Fig. 10. The numbers in the lower part of the schematic indicate the number of samples in each bin:



We suggest to include the schematic as a table in the appendix (Appendix B). We are adding the following explanation to our text to improve the clarity:

"These classes were defined such that the range of dynamic stability spanned by each bin is equally sized in the logarithmic space."

R2S25) Figure 10, caption: "bars" instead of "arrows".

We exchanged the "arrows" with "bars" in the figure caption, and also replaced all the other occurrences in the text.

R2S26) Figure 11: It is not clear which β model is represented here. Explain in the caption what is the grey zone.

Thanks for the comment. We are changing the caption as follows:

"This figure only presents results from REA model 3 (β_w). Left panel: β_w 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, bars 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. The grey area marks the +/- 10% range, which is the error assumed in EC applications."

R2S27) Page 20: "increasing z/L" is ambiguous since z/L could be either positive or negative

We are changing the wording to:

"with increasing (positive) z/L"

R2S28) Page 2; line 9: replace "sensible heat" and "latent heat" with "temperature" and "water vapour", respectively.

Thanks, we took this comment into account and changed the wording accordingly.

R2S29) Page 21: “methods. The diurnal course of the flux bias showed large deviations from the EC flux, particularly during transitions when the direction of the flux changed”. This is unclear, please rephrase.

Thanks for the hint. We are rephrasing the sentence as follows:

“However, during times of low proxy-scalar correlation, the variability of this ratio, measured by the RMSE, was large. This happened particularly at those times of the day when the direction (sign) of the flux changed.”

R2S30) Conclusion: A common name should be used for each model between the text and in Table 2.

We want to thank the reviewer for this comment, which will certainly improve the comprehensibility of the conclusions section. We are adding references to the model numbers 1-4, which are listed in Table 2, to the text.

R2S31) Section “Conclusions and practical recommendations”: The last sentence of the abstract contains a recommendation which is not present here.

Thank you for the hint - this recommendation should indeed be pointed out more directly! We propose to add the following to the Conclusions, and thus conclude the paper as follows:

“Based on the findings obtained in this study, we attempt to formulate the following general recommendations: For applications without deeper site-specific knowledge, we recommend using either the β_w or $\beta_{T, const}$ approach (model 3 or model 4). These two models have been shown to perform robustly and be less sensitive to changes in proxy-scalar similarity than model 1 and 2. In case of a well-known site, including scalar-scalar similarity, we propose to use the proxy-dependent approach in connection with a hyperbolic deadband (model 2). Model 2 yielded very similar results to model 1 with respect to the precision and accuracy measures considered in this study. However, hyperbolic deadbands are better suited to maximize the concentration difference between up- and downdraft reservoirs, which is of advantage when investigating fluxes of compounds with very low atmospheric concentrations.”