Authors response to reviews, paper "A multiyear eddy covariance and meteorological dataset from five pairs of agroforestry systems with open cropland or grassland in Northern Germany", submitted to *Earth System Science Data*, 10.5194/essd-2025-440

José Ángel Callejas-Rodelas[1], Justus van Ramshorst[1,2], Alexander Knohl[1,3], Lukas Siebicke[1], Dietmar Fellert[1], Marek Peksa[1], Dirk Böttger[4], and Christian Markwitz[1]

[1]University of Göttingen, Bioclimatology, Göttingen, Germany

[2]Quanterra Systems Ltd., Centenary House, Peninsula Park, Exeter EX2 7XE, UK

[3]Centre for Biodiversity and Land Use, University of Göttingen, Göttingen, Germany

[4]Soil Science of Tropical and Subtropical Ecosystems, University of Göttingen, Göttingen, Germany

The reviewers' comments are named as R1 (reviewer 1) and R2 (reviewer 2) followed by _C1, _C2, _C3, etc., numbering in order the comments. The authors' response is numbered in a similar way, using AR_C1, AR_C2, etc. The new figures crafted for this author's response are numbered AR1, AR2, etc., to distinguish them from the figures in the submitted manuscript.

1 Reviewer 1

1.1 General comments

R1_General comment. This dataset makes a valuable contribution to understanding the water–energy–carbon balance in agroforestry/cropland/grassland ecosystems and their potential as nature-based climate solutions. By providing 10 long-term, harmonized datasets across North Germany (78 site-years), the authors offer an important resource for the scientific community. The paper is well structured, and the authors clearly describe the measured variables, data processing, uncertainties, and comparisons with standardized datasets such as FLUXNET formatted ones. The paper also highlights well the ecological and social relevance of the work. I only have minor suggestions about data visualization, paper structure, and potential analyses as below:

AR_General comment. We appreciate the reviewer's comment about our manuscript. We are thankful for bringing out the main novelty of the study and key points, and also for the recommendations regarding changes that the manuscript should undergo. These major points are addressed throughout the comments in the following section.

1.2 Specific comments

R1_C1. In the method, the authors can consider merging Sections 2.3.2 and 2.3.3, since section 2.3.3 (XGBoost implementation) describes a gap-filling method that fits within section 2.3.2.

AR_C1. Thanks for the suggestion. Section 2.3.3 describes the general way XGBoost was implemented, both for meteorological and flux density data. That is why we decided to keep it as a separate section, because some information is common to both filling meteorological and eddy covariance time series. However, to make it slightly less confusing, we added this section as section 2.2.3, right after explaining the gap filling of meteorological data, and then referred back to it when explaining the gap filling of eddy covariance data. Then we added a reference to that subsection when explaining XGBoost specifically applied to eddy covariance ("The original code implementing XGBoost (see section 2.2.3) was used in ...").

R1_C2. It is surprised to see that measured SW_IN has no records in the last three years (Fig. 3), please clarify or explain this gap.

AR_C2. Thanks for the comment. This is not a gap of three years, it was a visualization effect. Most data were measured, but the filled data using ERA5-Land were drawn after the measured ones and they were overlapping. We reduced the size of the points, changed the colors for a better visualization, and drew the measured data at the end, so then it is clear they dominate. We still added a note to the caption of both figures 3 and 5, stating "Note that not all the filled points are visible in the figure; often, measured values impede the visualization of filled values.".

R1_C3. I am confused about Fig. 5. It is unclear to me whether XGBoost and REddyProc results overlap or correspond to different gap-filling periods. If they overlap, consider using or referring to a scatterplot for clarity. The same suggestion applies to Figure 3 if methods were applied over the same period.

AR_C3. Thanks for the comment. The data filled with REddyProc or XGBoost do not overlap, they correspond to different periods. We already used scatter plots for Figures 3 and 5, and the fact that points seem to overlap is just a visualization effect. However, we changed slightly the figure, same as for Figure 3, according to the previous comment AR_C2, to improve visualization. We added the same sentence to the caption of Figure 5.

R1_C4. Line 373 "For WS and WD, the correlation between measured and ERA5-Land data is worst and varies more across site" should include a figure reference.

AR_C4. Thanks for the suggestion. We included a reference to the figure in the Appendix: "For *WS* and *WD*, the correlation between measured and ERA5-Land data is worst (Fig. 5e and f)".

R1_C5. Line 392: Please provide supporting evidence or a reference for this statement, as it cannot be confirmed from Figure 5 alone.

AR_C5. Thanks for the comment. We added two references to this statement. In general, gradient boosting techniques, or tree ensemble algorithms, like XGBoost, smooth the extremes of the distributions due to the characteristics of the ensemble modeling implemented in the codes. This is discussed in some papers presenting these techniques, so we decided to cite the paper where XGBoost was presented (Chen and Guestrin, 2016) and another paper more generally talking about gradient boosting machines (Friedman, 2001). The new text reads now as follows: "However, XGBoost smoothed the extreme values of the distributions for all variables, as demonstrated by, for example, the gap-filled *LE* values, which were never below 0 (Fig. 5); this is a characteristic of gradient boosting techniques (Chen and Guestrin, 2016; Friedman, 2001)."

R1_C6. Line 572: Could the authors provide FC values with storage correction? Storage flux can bias FC measurements but can be estimated without installing a lower tower (see Pastorello et al., 2020). If tower height <3 m, storage flux ≈ 0; otherwise, refer to FLUXNET processing guidance.

AR_C6. Thanks for the suggestion. We calculated storage (*SC*) values using the single point measurements of the CO_2 concentration. We added *SC* to *FC* to get *NEE*, then applied the same processing routine with cleaning, filtering for the different *USTAR* thresholds, gap-filling and partitioning. We added the following paragraph to the section 2.3.1, "Processing of eddy covariance data":

> To obtain net ecosystem exchange (*NEE*), the storage term of CO_2 (*SC*) was calculated and added to *FC*: *NEE* = *FC* + *SC*. *SC* was calculated using the single point measurements of CO_2 concentration from the IRGA, since a vertical profile of the CO_2 mole fraction was not available at any of the stations. *SC* was calculated by taking the difference between each 30-min average of the CO_2 concentration (c_t), multiplied by the air molar density (ρ_m) and measurement height (z_m) and divided by the integration time (30 min), according to the following equation:
>
> $$SC = \int \rho_m \frac{\partial c}{\partial t} dz = \frac{\rho_m \cdot (c_t - c_{t-30min}) \cdot z_m}{30min} \tag{1}$$
>
> This is the procedure followed by FLUXNET (Pastorello et al., 2020) when a profile is non existent.
>
> After flux density calculations, *NEE*, *H* and *LE* were filtered to remove outliers and to ensure a good quality of the measurements. *FC* was also filtered to provide a clean version of it in the final datasets. Outliers were removed following the approach of Mauder et al. (2013), based on a median absolute deviation (MAD) filter, since it is a robust outlier detector (Leys et al.,

We updated the figures where we had displayed some *FC* values, to display *NEE*, concretely, Figures 5, 7 and C1. We updated the text explaining Fig. 7. The new results do not differ significantly from the previous ones, but there are some differences in the annual sums of *NEE* with respect to *FC* (Fig. 7). Nevertheless, we added the columns *SC*, cleaned *FC* and original *FC*, to the datasets, so then they are available to any user. We additionally changed the term *FC* by *NEE* where it corresponds throughout the text. Finally, we updated the corresponding paragraph in the discussion where the similarities and differences to the FLUXNET processing routine are discussed.

2 Reviewer 2

2.1 General comments

R2_General comment. Callejas-Rodelas et al. present a unique dataset of land-atmosphere fluxes of energy, water, and carbon for agroforestry systems and adjacent conventional systems. They use a novel low-cost eddy covariance technique and use state-of-the-art post-processing procedures. In my opinion, the dataset provides insights into ecosystems that are understudied given their potential for climate mitigation. The manuscript is well written and provides enough detail to understand the processing steps. My concern is that ancillary data is missing that is crucial for the interpretation of the fluxes to assess climate mitigation potential. First, without information on standing biomass, carbon export through harvest, and carbon import through organic fertiliser, the existing flux data will be difficult to interpret. The authors provide a table with information on crop rotation, but I believe this is insufficient. At the minimum, the authors should discuss how this issue can be addressed in studies using this dataset. Second, the authors mention the issue of tower location dependency of flux measurements. This is an important issue and should be addressed in more detailed. They provide yearly flux footprint and highlight the need for additional flux footprint modelling. However, to maximise the insights gained from footprint models, the authors should consider publishing spatial maps/data on land cover and properties surrounding the flux towers. End users can then use these maps to conduct their own footprint analyses.

AR_General comment. We appreciate the reviewer's comment about our manuscript. Indeed, carbon exports are important to assess the longer term impact of AF on carbon balance. We added a table in the Appendix, with the type of data and corresponding DOIs where the data on carbon yield from crops and straw can be found (see reply AR_C4). The readers can then refer to this table and the data within the repositories linked with the DOIs. Furthermore, in further analyses we plan to consider as well all these data for better accounting for carbon exports through harvest. We clarified in the caption of Figure 7 that the annual carbon sums do not include these data and should be considered carefully. Regarding the land cover maps around the stations, we included shape files (.shp) with the basic land cover information for all sites in the repository. We added a paragraph explaining this in section 5, "Data records and usage of the dataset", that reads as follows: " Furthermore, land cover maps containing the distinction between tree and crop areas at the sites were included in the repository. The maps are in the format .qgz, which can be opened with any tool for geospatial analysis. At Wendhausen and Forst, the different crops at the footprint area of the stations were also outlined, and the corresponding information to the crops rotation can be found in Table A1".

2.2 SPECIFIC COMMENTS

R2_C1. Line 8: How do the climatic regions differ?

AR_C1. According to the Köppen-Geiger classification, the regions are slightly different, with two sites belonging to the category of temperate oceanic climate (Cfb), Mariensee and Vechta, and the three other sites belonging to the category of continental climate with warm summers and no dry season (Dfb). We added a column to Table 1 with the climatic zone. Even though some of the sites share the same climate zone, there were differences in the patterns and seasonal behavior of the meteorological variables across the measured period. However, for clarity, we removed the word "climatic" from that sentence in the abstract, so the sentence now reads "... in five distinct regions of Germany".

R2_C2. Table 1: Why not using the reference period 1991-2020?

AR_C2. The reference period was taken as 1981-2010 to exclude project years from the average, so then a better comparison on the current conditions is done with respect to near-past conditions.

R2_C3. Table 1: The impact from the different soil properties on fluxes could also be discussed. Mariensee stands out with a very high soil organic carbon content.

AR_C3. We appreciate the suggestion. However, since this is a data description paper, we will leave most discussions related to ecosystem functioning and differences between AF, OC or OG, and between the studied regions, for future work. Currently we are preparing another publication were the carbon fluxes are discussed in detail.

R2_C4. Line 89: Trees were harvested but information on exported biomass/carbon is missing.

AR_C4. As written above in response to the main comment, indeed, carbon exports are important to assess the longer term impact of AF on carbon balance. We added a table in the Appendix, with the type of data and corresponding DOIs where the data on carbon yield from crops and straw can be found. The readers can then refer to this table and the data within the repositories linked with the DOIs. This table looks like this:

Table A3. References and DOIs for datasets related to biomass and carbon yield data for trees, crops and grasses at the project sites.

	Reference	DOI
Dornburg	Choe et al. (2025a)	https://doi.org/10.20387/BONARES-BKJE-VB72
Forst	Choe et al. (2025b)	(preprint)
Mariensee	Langhof and Swieter (2024)	https://doi.org/10.1007/s10457-024-00963-2
Vechta	Choe et al. (2025a)	https://doi.org/10.20387/BONARES-BKJE-VB72
Wendhausen	Swieter and Langhof (2020)	https://doi.org/10.20387/bonares-agdc-m0dm

Furthermore, in future analyses we plan to consider as well all these data for better accounting for carbon exports through harvest. We clarified in the caption of Figure 7 that the annual carbon sums do not include these data and should be considered carefully, adding the following sentences: "Note that these sums do not included the carbon exports through harvest of crops and trees. Comprehensive analyses on carbon balance over the studied sites should include those data, accessible through DOIs in Table A3.".

R2_C5. Figure 2: The target area for Mariensee is much smaller than the flux footprint. How useful are the flux measurements then to understand the agroforestry impact? Additionally, the flux footprint overlap and fluxes are thus not independent between AF and OC/OG.

AR_C5. Thanks for the comment. This is indeed one of the main shortcomings of this dataset for Mariensee, however its impact of flux measurements remains uncertain. Some footprint filtering could be applied to flux data, but this would remove even more data from the measurements, which means more data need to be filled; and, additionally, the fact that footprint climatology overlap does not mean that all the flux contributions from a specific period are the same for both stations. The major contributions come from around the stations, in a relatively small area, with decreasing importance of every patch of land with increasing distance to the tower. A more detailed footprint modeling would need to be based on a proper turbulence field characterization, prior to use the footprint model after Kljun et al. (2015) or another one similar to it, including the detailed land cover information around the stations. This is planned for future work, but it was not included in this dataset preparation since it would itself be part of a completely different study. The main objective of this paper was to present and describe this dataset, and at the same time provide a basis for further studies on this and other topics. In relation to this comment and the major comment, we added to the repository the land cover maps of the different sites, with the distinctions between crops and trees. Also, we added the additional crops that were cultivated in Wendhausen and Forst, besides the main crop in the area. The maps, added as .qgz files, can be opened with any tool for geospatial analysis. We added the corresponding information about the maps within the section 5, "Data records and usage of the dataset". We added the following paragraph: "Furthermore, land cover maps containing the distinction between tree and crop areas at the sites were included in the repository. The maps are in the format .qgz, which can be opened with any tool for geospatial analysis. At Wendhausen and Forst, the different crops at the footprint area of the stations were also outlined, and the corresponding information to the crops rotation can be found in Table A1."

R2_C6. Line 183: For testing the gap-filling performance (and particularly for testing XGBoost), how was the dataset split in training and test? Was it done randomly or blocked? Due to the high autocorrelation in eddy covariance data, random selection usually results in much higher correlation and lower RMSE.

AR_C6. The split was done randomly, not blocked. However, the cross validation implemented in the code was introduced to avoid overfitting and to get an average of the performance of the model in many different test datasets. Therefore, we consider the RMSE estimate as reliable.

R2_C7. Line 289: It is important to establish if difference between lower-cost and EC methods are truly random or if a systematic bias exists. This information is important to better understand the nature of the calculated error.

AR_C7. According to the previous validation of these lower-cost (LC-EC) setups, published in Callejas-Rodelas et al. (2024) and van Ramshorst et al. (2024), the LC-EC suffer from an intrinsic larger spectral attenuation, which is considered a systematic error. However, when we assigned an individual error to the 30-min measurements performed with LC-EC setups, the nature of that error is random because we did not explicitly assign a direction to it. That means, the error is an uncertainty range around the exact value measured by the LC-EC, and that is why, when propagating it through a sum, its magnitude reduces with the length of the time series used for the sum aggregation. This method was applied in Callejas-Rodelas et al. (2025a), as briefly explained in the current paper.

To provide a more robust estimate, we applied a different method for uncertainty estimates, based on Richardson et al. (2007). We split the previous section 2.3.4 in two, now new section 2.3.3 is "Evaluation of gap-filling of eddy covariance data" and new section 2.3.4 is "Uncertainty estimates of eddy covariance data". This is how the new text looks:

2.3.3 Evaluation of gap-filling of eddy covariance data

Both REddyProc and XGBoost reproduced the diel cycle of measured data, for *NEE* (Fig. C1a) and *LE* (Fig. C1b). The gap-filling led to a reduction of the magnitude of the measured values, due to the fact that most gaps were present during winter and during nighttime, with smaller flux density magnitudes. However, the similarity in diel cycles between measured and filled data can be considered an indicator of reliability of the gap-filling procedure, besides other sensitivity or uncertainty analysis that may be performed.

Furthermore, RMSE was evaluated for the application of XGBoost across different scenarios of training and testing data. The RMSE values displayed in Table 5 were generally small and indicate a good performance of the technique.

14

2.3.4 Uncertainty estimates of eddy covariance data

A method to evaluate the uncertainty in aggregates of eddy covariance data (such as annual carbon or water balances) was proposed in Callejas-Rodelas et al. (2025a). This method consisted in assigning individual errors to the 30-min data in two distinct manners. In the first case, errors were calculated as the deviation of fluxes between lower-cost and conventional EC setups (Callejas-Rodelas et al., 2024 and van Ramshorst et al., 2024). This deviation was considered as the worst case RMSE of the linear regression models of *NEE* and *LE*, respectively. In the second case, errors were calculated as the sum of that deviation and the random error (Callejas-Rodelas et al., 2025a). Errors in gap-filled data with REddyProc were defined as the standard deviation of the data points used for gap-filling (Wutzler et al., 2018). Finally, the error in gap-filled data with XGBoost was taken as the RMSE between modeled and measured data, after fitting the XGBoost regressor and prior to predicting all missing data, similarly to the meteorological data filled with XGBoost. RMSE values of gap-filled *NEE*, *LE* and *H* are displayed in Table 5. The individual errors can be propagated if a cumulative sum is calculated, as explained in Section 2.5 of the paper by Callejas-Rodelas et al. (2025a). The proposed method leads to small uncertainty estimates for very long time series, as the propagated error reduces its magnitude with the square root of the number of data. This is due to the nature of the error, which
by Callejas-Rodelas et al. (2025a). The proposed method leads to small uncertainty estimates for very long time series, as the propagated error reduces its magnitude with the square root of the number of data. This is due to the nature of the error, which
is considered fully random: when an uncertainty range is assigned to a single 30-min data point, the underlying assumption is that the data is equal to its value $\pm$ the uncertainty. This method, therefore, is not suitable for considering systematic deviations between setups with a known performance, such as our LC-EC and CON-EC setups.

For this reason, we applied a different method to evaluate the uncertainty in annual or multi-annual sums of *NEE* or *ET*. This method was based on Richardson and Hollinger (2007) with slight modifications. For this error evaluation, we used the filtered and cleaned data corresponding to the *USTAR* 50 % scenario for all sites. The method, as described in Richardson and Hollinger (2007), consisted in two distinct steps.

First, we generated 50 different synthetic datasets for *NEE* and *LE*, using neural networks implemented through the library *keras* (Chollet et al., 2015) implemented in the *tensorflow* platform (**?**). We added random noise to the time series of *NEE* and *LE*, so then each of the 50 datasets had different levels of noise. The code to generate the synthetic datasets was based on the code by Vekuri et al. (2023) for their evaluation of XGBoost against REddyProc. We then added the same gaps existing in the original time series to the synthetic datasets, and applied the same processing routine described previously: first, short gaps were filled with REddyProc, then longer gaps were filled with XGBoost. We calculated the standard deviation across the 50 different multi-annual sums for *NEE* (σ_{R_NEE}) and *LE* (σ_{R_LE}) to get the uncertainty due to the random noise in the original data.

Second, we generated a single gap-free synthetic dataset and created different short- and long-gap scenarios, selecting only one year of data. For all sites we used 2020, because it was the most complete year, except for Wendhausen OC for which we used 2022 as 2020 was not available. Short gaps were added covering randomly 30 % of the dataset. Long gaps were added in

Second, we generated a single gap-free synthetic dataset and created different short- and long-gap scenarios, selecting only one year of data. For all sites we used 2020, because it was the most complete year, except for Wendhausen OC for which we used 2022 as 2020 was not available. Short gaps were added covering randomly 30 % of the dataset. Long gaps were added in addition to the short gaps, starting in day one of the year, until day 365, in intervals of 1 to 28 days, increasing in steps of 3 days, with the gap starting in day one at first, then increasing in intervals of 10 days until day 361. This means for each station, we created 370 short-gap scenarios, and 370 long-gap scenarios. All these scenarios were filled using the same routine as explained

previously. Then, we calculated the difference in annual sums of *NEE* ($\Delta NEE_{j,k}$) and *LE* ($\Delta LE_{j,k}$) between the short- and long-gap scenario. This difference was the error due to the presence of a gap of length *j* starting on day *k*. This was calculated for all j, k scenarios; then, we divided the year in 12 months (periods m), and calculated for each combination of j and m the standard deviation of $\Delta NEE_{j,k}$ or $\Delta LE_{j,k}$ across all k for each month m across all k scenarios (the long-gap scenarios) to get $\sigma_m(\Delta NEE_j)$ and $\sigma_m(\Delta LE_j)$, respectively). After this, we calculated the slope of the relation between gap length (*j*) and $\sigma_m(\Delta NEE_j)$ or $\sigma_m(\Delta LE_j)$. This slope was used to get the uncertainty due to long gaps ($\sigma_{\mathrm{LG}}(NEE)$ or $\sigma_{\mathrm{LG}}(LE)$) by multiplying it by the gap length in days. Finally, we calculated the error in the annual sums by adding in quadrature the error due to noise and the error due to long gaps, both for *NEE* and *LE*:

$$\sigma_{\mathrm{TOT}} = \sqrt{\sigma_{\mathrm{R}}^2 + \sigma_{\mathrm{LG}}^2} \qquad (2)$$

Finally, we added in quadrature the previously described error corresponding to long gaps, and the error described in the previous paragraph corresponding to random noise. This was considered the total error in the final sum. The advantage of this method is that it can be applied to any annual or multi-annual sum, by just selecting the corresponding period in the short-gap evaluation, and using the relations between gap length and standard deviation for different months of the year, corresponding to the long-gap evaluation. On the other hand, there exist two main hindrances. First, it could lead to bias in the error evaluation due to the nature of these ecosystems: since crops rotate year to year, therefore assuming similar uncertainty across the season might be wrong. Second, this method does not evaluate the performance of the gap-filling itself, so it assumes no uncertainty related to the method selection, just due to the noise and gaps in the measured time series. However, in general it provides a more robust estimate of the uncertainty compared to the simple error propagation described at the beginning of this section.

Within section 2.3.5, we mention briefly the method based on error propagation explained in the previous paper by Callejas-Rodelas et al. (2025a), and explain its limitation since it considers the error due to the lower-cost eddy covariance as random, and when it propagates through a sum its magnitude reduces with the length of the time series used for the sum. The remaining of section 2.3.5 focuses on explaining the method of Richardson et al. (2007) with the slight modifications we implemented. Thanks to this method implementation, we now could provide actual numbers to the uncertainty estimates. We run such method on the *USTAR* 50 % threshold scenario, and we added a Table to section 3.2, "Eddy covariance flux densities", after Figure 7. The Table contains annual sums of *ET* and *NEE*, with the uncertainty values obtained through that method, and a paragraph was added explaining the content of the Table. This Table looks now like this:

Table 8. Annual sums of *ET* and *NEE*, corresponding to the *USTAR* 50 % scenario, together with the uncertainty estimates based on Richardson and Hollinger (2007) (see section 2.3.4). The uncertainty in *ET* was originally calculated for *LE*, then the error in *ET* was taken based on the same ratio between the error in *LE* and the annual sum of *LE*.

	$ET \pm \sigma_{\mathrm{TOT}}(ET)$ (mm)						$NEE \pm \sigma_{\mathrm{TOT}}(NEE)$ (g C m^{-2})					
	2019	2020	2021	2022	2023	2024	2019	2020	2021	2022	2023	2024
Dornburg AF	128.12 ± 5.70	482.54 ± 7.97	542.88 ± 5.72	531.36 ± 19.95	795.56 ± 15.38	610.89 ± 27.40	98.40 ± 82.64	-10.11 ± 46.06	-655.00 ± 29.06	-84.49 ± 55.94	-86.24 ± 51.59	-416.76 ± 73.19
Dornburg OC	198.06 ± 4.37	528.79 ± 6.02	—	433.54 ± 13.84	535.94 ± 5.26	401.53 ± 13.49	289.99 ± 56.53	192.06 ± 50.98	—	-142.09 ± 46.05	21.37 ± 30.82	-615.56 ± 67.80
Forst AF	184.14 ± 3.83	369.06 ± 11.65	392.52 ± 5.67	456.03 ± 5.79	733.26 ± 8.92	676.06 ± 28.87	60.65 ± 14.78	-128.88 ± 67.66	-292.52 ± 28.70	-264.44 ± 23.69	-296.14 ± 44.40	-155.58 ± 48.24
Forst OC	224.42 ± 4.45	379.36 ± 7.69	415.80 ± 9.56	381.07 ± 6.57	550.92 ± 19.08	704.73 ± 10.81	-85.93 ± 15.41	-42.46 ± 21.78	-195.23 ± 28.46	-18.00 ± 19.12	-62.86 ± 30.25	26.05 ± 29.72
Mariensee AF	377.05 ± 11.41	506.49 ± 7.47	595.87 ± 13.89	—	—	—	-153.23 ± 52.27	-152.14 ± 32.98	8.60 ± 85.65	—	—	—
Mariensee OC	317.95 ± 31.69	405.62 ± 12.84	631.33 ± 11.25	—	—	—	333.96 ± 56.42	354.77 ± 35.97	111.81 ± 17.53	—	—	—
Vechta AF	134.53 ± 9.06	572.59 ± 31.80	544.36 ± 36.08	—	—	—	207.15 ± 42.46	190.63 ± 75.34	405.86 ± 96.75	—	—	—
Vechta OC	125.84 ± 2.14	387.82 ± 4.51	387.85 ± 9.92	425.04 ± 3.54	372.99 ± 8.66	485.96 ± 4.68	89.24 ± 15.32	72.64 ± 31.23	-73.31 ± 66.52	97.49 ± 27.01	329.31 ± 93.25	-98.62 ± 53.92
Wendhausen AF	333.31 ± 23.60	—	410.21 ± 5.06	425.48 ± 7.44	688.02 ± 8.48	487.36 ± 10.58	-530.88 ± 100.34	—	229.44 ± 66.64	167.73 ± 80.30	90.60 ± 52.03	189.50 ± 57.47
Wendhausen OC	349.21 ± 27.53	—	443.79 ± 5.85	503.01 ± 7.40	592.11 ± 9.84	740.27 ± 12.68	-182.13 ± 119.95	—	-18.72 ± 28.38	252.12 ± 28.02	-108.59 ± 49.65	328.40 ± 52.19

Please refer to the new version of the paper and the track-changes version to see all these changes.

R2_C8. Line 300: Here, and throughout the manuscript, it would be helpful to quantitatively support the main text with actual numbers. As it is written now, most statements remain qualitative.

AR_C8. We changed this section, according to what was described in the previous answer (AR_C7). Please refer to AR_C7 for clarification.

R2_C9. Line 308: Why did the daytime partitioning fail? This information would be useful.

AR_C9. We were not sure about the error, since the error message was that some of the time series used to partition were not varying. However, this changed after the new processing including storage terms. The daytime partitioning failed only for the last few *USTAR* scenarios, from percentile 87.5 to percentile 97.5. It could be due to not enough data being available full filling the good quality requirement of REddyProc (QC=0), because for the top *USTAR* percentiles more data are removed from the original time series. We updated the corresponding line in the text: " There was one exception, for Wendhausen OC, where the daytime partitioning method failed for the upper percentiles (87.75 to 97.5) and only nighttime *GPP* and *RECO* were provided for those *USTAR* scenarios."

R2_C10. Table 6: I am not sure how meaningful the mean parameters for all variables other than zm, z, and ha are.

AR_C10. The objective was to provide some mean characteristics of the parameters used in the footprint modeling, but indeed it is not something very useful as we already provide all the parameters as time series anyways. We removed all unnecessary columns and left only zm, z, h_a and $USTAR$.

R2_C11. Table 7: I would suggest to only compare meteorological conditions between sites for overlapping periods.

AR_C11. We thank the reviewer for the suggestion. We re-calculated the averages for similar periods, except for the stations that covered less years. We did not include though similar periods for all sites because this would miss many years from all the stations, as measurements started in Vechta in 2019 and stopped in Mariensee in 2021, and the idea was to provide a climatic characterization of the sites without looking in detail at the patterns, variability or differences across sites and AF or OC/OG. We clarified all of this in the caption so then it is not misleading and we updated the values in the table. The new caption now reads: "Averages of shortwave incoming radiation (*SW_IN*, W m^{-2}), atmospheric pressure (*PA*, kPa), air temperature (*TA*, °C), relative humidity (*RH*, %), vapor pressure deficit (*VPD*, kPa), precipitation (*P*, mm), wind speed (*WS*, m s^{-1}) and wind direction (*WD*, °) for all sites. Averages were calculated across the period 1 July 2016 to 1 July 2024 for Dornburg, Forst and Wendhausen; 1 July 2016 to 1 July 2021 for Mariensee; and 1 July 2019 to 1 July 2024 for Vechta. For *P*, the value was calculated as the average of the annual cumulative sums across the corresponding period. For the other variables, the average was calculated as the average of all 30 min values across the corresponding period. *WD* is expressed in degrees (°) from the North direction."

R2_C12. Figure 3: It looks like the wind speed is decreasing over time. Is that due to the growing trees? These are important findings and should be discussed.

AR_C12. We did not analyze in detail trends in wind speed but it could be indeed due to the growing trees. However, it could also be an effect of some distinct climatic patterns and year-to-year variability. We added a sentence to the corresponding paragraph " *WS* slightly reduced in magnitude over time, mostly after 2021, an effect that could be attributed to the increasing tree height after the tree harvest in that year at the site.", to clarify this. We decided not to add more details because the focus of the paper is mainly to describe the dataset processing and some of its major features.

R2_C13. Line 420: I would suggest to not report partial years.

AR_C13. We understand that full years are most valuable for future data users. Given, however, that we don't oversee all potential applications, e.g. just comparing certain months, in future work, we prefer to provide all data we have as part of the paper and then let future users decide which data to select.

References

Callejas-Rodelas, J. Á., Knohl, A., van Ramshorst, J., Mammarella, I., and Markwitz, C.: Comparison between Lower-Cost and Conventional Eddy Covariance Setups for CO_2 and Evapotranspiration Measurements above Monocropping and Agroforestry Systems, Agricultural and Forest Meteorology, 354, 110086, https://doi.org/10.1016/j.agrformet.2024.110086, 2024.

Callejas-Rodelas, J. Á., Knohl, A., Mammarella, I., Vesala, T., Peltola, O., and Markwitz, C.: Does Increased Spatial Replication above Heterogeneous Agroforestry Improve the Representativeness of Eddy Covariance Measurements?, Biogeosciences, 22, 4507–4529, https://doi.org/10.5194/bg-22-4507-2025, 2025a.

Chu, H., Baldocchi, D. D., Poindexter, C., Abraha, M., Desai, A. R., Bohrer, G., Arain, M. A., Griffis, T., Blanken, P. D., O'Halloran, T. L., Thomas, R. Q., Zhang, Q., Burns, S. P., Frank, J. M., Christian, D., Brown, S., Black, T. A., Gough, C. M., Law, B. E., Lee, X., Chen, J., Reed, D. E., Massman, W. J., Clark, K., Hatfield, J., Prueger, J., Bracho, R., Baker, J. M., and Martin, T. A.: Temporal Dynamics of Aerodynamic Canopy Height Derived From Eddy Covariance Momentum Flux Data Across North American Flux Networks, Geophysical Research Letters, 45, 9275–9287, https://doi.org/10.1029/2018GL079306, 2018.

Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A Simple Two-Dimensional Parameterisation for Flux Footprint Prediction (FFP), Geoscientific Model Development, 8, 3695–3713, https://doi.org/10.5194/gmd-8-3695-2015, 2015.

Richardson, A. D. and Hollinger, D. Y.: A Method to Estimate the Additional Uncertainty in Gap-Filled NEE Resulting from Long Gaps in the CO_2 Flux Record, Agricultural and Forest Meteorology, 147, 199–208, https://doi.org/10.1016/j.agrformet.2007.06.004, 2007.

van Ramshorst, J. G. V., Knohl, A., Callejas-Rodelas, J. Á, Clement, R., Hill, T. C., Siebicke, L., and Markwitz, C.: Lower-Cost Eddy Covariance for CO_2 and H_2O Fluxes over Grassland and Agroforestry, https:

//doi.org/10.5194/amt-2024-30, 2024.