Dear Dr. Metzger,

We want to sincerely thank you for the thoughtful review and constructive feedback you provided. We very much appreciate the time you have taken, and we are pleased that you found the study highly relevant to the biogeosciences community. Your detailed insights and recommendations are very valuable to us - below, we provide point-by-point responses with our answers, reflections and adjustments.

We believe the paper improved thanks to your input.

Laura van der Poel and co-authors

1. Figure 2

Please add a scale bar to clarify the geographic extent of the EC tower network and flight tracks (i.e., are we looking at 10s, 100s, or 1,000s of km?).

We added a scale bar and changed the figure to the following:

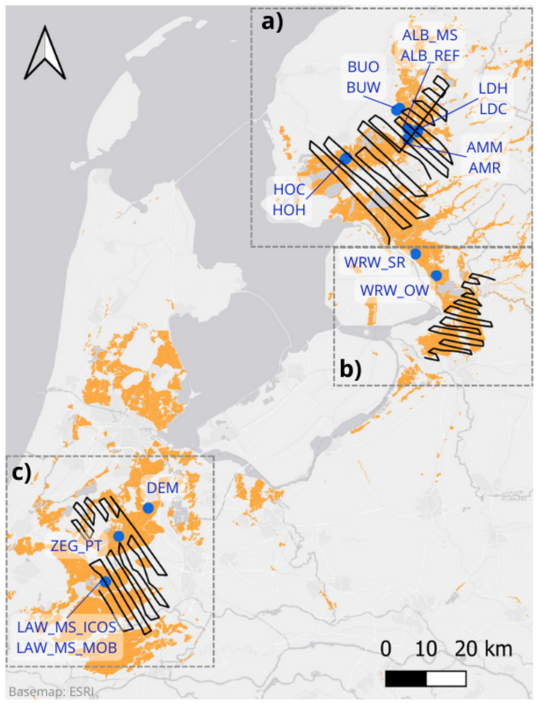

Figure 2: The EC tower network used in this study, with three flight tracks over the study areas: a) Friesland, b) Overijssel and c) Groene Hart. Peat distribution is shown in orange. Airborne fluxes are calculated over the entire flight tracks, including the turn, since the banking angle was kept less than 15 degrees. General information on the EC sites and processing can be found in 2.1.3 and Appendix A. For more detailed, site-specific information, see Kruijt et al. (2023). The map was made in QGIS using an ESRI base map; peat distribution was obtained from the soil map (Wageningen Environmental Research, 2024).

2. Figure 3

The ZEG_PT tower is hard to locate even in the inset. Consider increasing font size or adding an arrow pointing to it.

We adjusted the figure such that readability of ZEG_PT is improved. Furthermore, we cleaned up the figure for clarity and extended the caption:

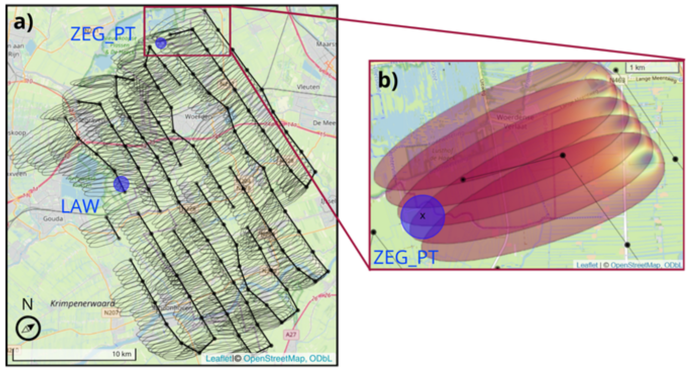

Figure 3: Airborne sub-footprints of a typical flight over the Groene Hart, with flight altitude of 60 meters. In this study, we used static circular footprints for towers, which are shown in blue for LAW and ZEG_PT. In **a)**, five sub-footprints are visible for every 2-km window, where the contour lines represent the area from which 80% of the measured flux originates. All sub-footprints were overlaid with spatial data, and subsequently combined and normalized to get the final footprint values. The wind-rose shows the average wind direction. In **b)**, we show the contribution distribution within the first five sub-footprints: blue indicates the highest contribution, red indicates the lowest. The black 'x' denotes the ZEG_PT tower location. In both **a)** and **b)**, the differences in tower and airborne footprints are visualized.

3. Figure 4

Airborne fluxes seem to show lower amplitude compared to tower fluxes. Have the authors considered vertical flux divergence or other systematic differences in data capture and processing? A cluster analysis could help evaluate whether the two platforms yield comparable flux regimes after controlling for diurnal cycle and differing surface heterogeneity in the footprints. Please also see comment 13 below.

Thanks for raising this issue. Referring to figure 4 we do not concur that airborne fluxes *systematically* seem to show lower amplitude compared to tower fluxes. For most of the 16 months shown here, there are sites with higher as well as lower magnitude fluxes, not only in their median, but also their interquartile ranges do not overlap. Only in 1 out of 16 months (oct21) the airborne fluxes are lower than any of the sites. We have not done a formal cluster analysis.

There is more to be said about flux divergence. First, we aim to minimize this by flying nominally at 200 ft/60m above the surface. This may be considered to be in the 'constant flux layer' given that over all flights the average boundary layer height according to ERA5 re-analysis (for a grid box centered on each of the 3 regions) is 870m at the time of the flights.

Few studies have been done on CO_2 flux divergence: de Arellano et al. (2004) (Cabauw, in the Groene Hart area studied here); Casso-Torralba et al. (2008) (Cabauw again); Vellinga et al. (2010) (Supplementary Material, SW France). CO_2 flux divergence above typical observation heights of towers is notoriously hard to quantify due to transient effects and possibly missed fluxes from high and low frequency contributions. Further it is complicated by the entrainment flux which for CO_2 can be significant and even larger than surface flux in the morning, when night time high-concentration boundary layers are flushed with low concentration air from the free troposphere. In the afternoon the entrainment flux may even reverse sign, as boundary layer concentrations may become lower than free troposheric values, due to strong uptake (photosynthesis). Thus, also the sign of the flux divergence is not a given (as for the vapor flux). Such complications prohibit assessments of CO_2 flux divergence without dedicated observation strategies, let alone allow simple corrections. Finally, neglecting advection, flux divergence equals the scalar storage term, i.e. temporal change in CO_2 concentration. From our tower observations we know these to be small around mid-day. For all these reasons, with Vellinga et al. (2010) and e.g. Meesters et al. (2012) we assume CO_2 flux divergence errors are arguably smaller than other errors, not of constant sign (so partially cancel out) and therefore they are neglected here.

We added a few sentences in the discussion of Figure 4 as well as in the discussion.

4. Table 2

Please include EC tower measurement heights in the text to facilitate comparisons with remote sensing data resolutions. For example, a 250 m MODIS resolution would match a 250 m tall tower (e.g., Xu et al., 2017) [https://doi.org/10.1016/j.agrformet.2016.07.019], whereas shorter towers may be better represented by higher-resolution products such as Sentinel.

All flux towers are relatively low, with measurement heights ranging from 1.5 to 6 m. The aircraft flies at 60 m height (200ft by special permission of aviation authorities), which is substantially lower than the tall tower by Xu et al. (2017). Given these relatively low observation heights, the higher spatial resolution of Sentinel products may offer improvements for future studies compared to MODIS. We mention potential for Sentinel in Section 4.5.1. and added a few sentences on the towers in Section 2.1.3:

"All sites were equipped with open-path gas analyzers, except LAW_MS_ICOS, which used a closed-path CO_2 sensor. Otherwise, sensor set ups were identical. The equipment was mounted on top of telescopic masts and arranged perpendicular to prevailing southwest winds. Measuring height ranged between 1.5 m and 6 m, based on the desired footprint size."

We also specified this in the caption of Table 2: "Overview of supplementary data sources. For comparison, tower heights ranged between 1.5 and 6 meters, with footprints spanning several hundred meters; airborne flying altitude is 60 meters, with footprints spanning several kilometers."

5. Table 3

The regional model explains 60% of the variance, which is not dramatically higher than global models (e.g., Jung et al., 2020 [https://www.biogeosciences.net/17/1343/2020/]). Please discuss this apparent incongruity, given the expected higher information density at the regional scale.

Thank you for the comment. We added a few sentences discussing this apparent incongruity in Section 4.1, line 336: "Still, the R^2 is not substantially higher than that of global models (e.g., Jung et al. (2020)), despite the higher information density in our study. We attribute this to the relatively subtle variability within our study area, which encompasses seemingly similar systems in terms of land use, climate, and flux characteristics - making it more difficult to distinguish patterns than at the global scale."

6. Figure 5

The comparison between regional and local models is insightful but largely confirms expectations. Consider moving the detailed discussion and figures to the appendix and condensing the main text.

We considered your suggestion. However, we believe the figure adds to the main text as it validates our choice of using the model based on all areas. Furthermore, it shows the large difference in the performance of the Overijssel model when predicting for other areas.

7. Figure 6

Clarify the sign convention for water table depth (e.g., lower water tables as positive depth vs. negative height). The current color coding and SHAP interpretation may confuse readers without a consistent definition.

We added units for every feature in the figure and extended the figure caption, to the following:

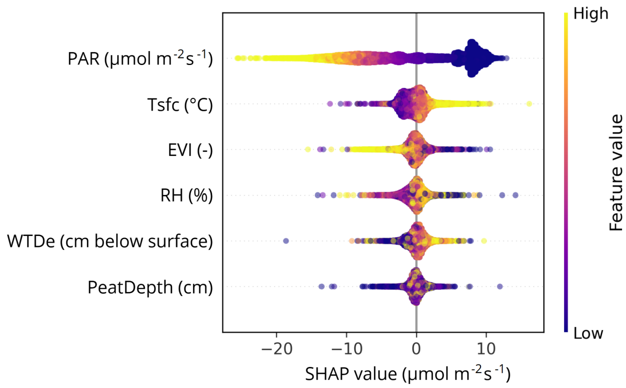

Figure 6: SHAP values for all features in the final model, representing the individual contribution of each feature to the final model outcome, depending on the value of that feature. The thickness indicates the amount of data points. For example, there are many data points with low PAR and a positive SHAP values, indicating that the model assigns a positive contribution to predicted CO_2 flux under low-light (i.e. nighttime) conditions. Abbreviations: PAR (photosynthetically active radiation), Tsfc (surface temperature), EVI (enhanced vegetation index), RH (relative humidity), WTDe (effective water table depth).

8. Figure 7

This figure helps resolve the confusion in Figure 6 by explicitly stating water table depth in cm below ground. Consider using negative values (e.g., -150 to 0 cm) throughout for better intuitiveness also for non-subject matter experts.

We ensured consistency throughout the text and figures, writing explicitly that WTD is in (centi)meters below the surface. We keep this convention, as this is commonly done by groundwater-CO_2 related studies (Aben et al., 2024; Boonman et al., 2022; Evans et al., 2021; Koch et al., 2023).

9. Figure 10

This simulation result is the centerpiece of the manuscript. Please revise the phrase ”increased with 10 cm” to ”increased by 10 cm” for clarity.

Thank you for pointing this out. We changed into “increased by 10 cm”.

10. Figure 11

The SHAP-derived functional relationships are particularly powerful for integrating direct measurements and non-linear interactions. Consider extending the outlook to include the potential for assessing CH4 fluxes and albedo effects using the same EC framework, enabling full net radiative forcing (NRF) or CO2e tradeoff evaluations through a consistent methodology.

Thank you for the suggestion. While we think this would be an interesting extension of the study, we doubt whether this is feasible due to the following considerations:

- We do not have airborne fluxes of CH_4, only concentrations.
- CH_4 fluxes are primarily driven by soil temperature and groundwater level, rather than radiation (Buzacott et al., 2024), resulting in less temporal (daily) variability and making them more difficult to model with a ML approach as used in our study.
- CH_4 flux data are subject to stricter quality filtering compared to CO_2, which reduces the data available for training.
- Achieving a full NRF balance would additionally require consideration of N_2O emissions, a significant term given the high cattle densities in the Netherlands, which are costly and challenging to measure, especially from the aircraft.

Considering all, we did not add the potential for assessing CH_4 fluxes to the outlook.

11. Line 130

Clarify whether the ”2 km windows” were applied in a moving window fashion, and if so, specify the step size, degree of overlap, and the import of those choices on downstream data integration and results.

The 2 km windows were not applied in a moving window fashion: there was no overlap, as to our opinion the overlap would have compromised the independence of the sequential measurements, with possible unwanted effects in the ML models. We now clarified this on line 130: “The fluxes at these two scales are then summed in non-overlapping 2 km windows to get the flux over all scales.”

12. Line 138

Vaughan et al. (2021) and Serafimovich et al. (2018) used the Metzger et al. (2012) [https://doi.org/10.5194/amt-5-1699-2012] footprint model, not Kljun et al. (2015). Please correct.

As the sentence on line 138 describes examples of the Kljun et al. (2015) model, we removed Vaughan et al. (2021) and Serafimovich et al. (2018).

13. Line 178

Please describe how EC tower data processing (e.g., block averaging, de-spiking, spectral correction, density correction, quality flags) compares with airborne EC data processing (e.g., Wavelet decomposition). This will help readers assess the interoperability of the datasets. Please also see comment 3 above.

We added a table in the Appendix with all processing steps for tower and airborne data.

Table 1: Processing steps for tower and airborne flux calculation and filtering. EddyPro was used for tower fluxes.

Processing step	**Tower**	**Airborne**
	Flux calculation	
Block-averaging	✓(30 min)	✓(2 km)
Wavelet decomposition	-	✓
Reynolds decomposition	✓	-
WPL[1] correction	✓	✓
Frequency loss correction	✓	High freq. not needed because at operating altitude flxues are carried by eddies <10Hz or 4m. Low freq. not needed because of wavelet decomposition.
Tilt correction radiation sensors[1]	-	✓
	Filters	
Signal strength filtering	✓Remove Received Signal Strength Indicator (RSSI) < 70 (McDermitt et al., 2011)	-
Precipitation	✓Remove timesteps with precipitation (sensor performance affected)	(no flights with precipitation)
Wind direction	✓Exclude fluxes from undesired wind sectors (site-specific)	-
u* filtering	✓Remove stable night data below u* threshold (site-specific, moving point test by Papale et al. (2006))	
Stationarity and ITC[3] test	✓	✓
Meteorological measurements: physical range filter	✓	✓
Flux magnitude	$-100 < CO_2$ (μmol $m^{-2}s^{-1}$) < 100	$-50 < CO_2$ (μmol $m^{-2}s^{-1}$) < 50
Quality flags	✓All hard flags by Vickers and Mahrt (1997). ✓Keep only quality flag = 2 (Foken et al., 2004).	✓Keep quality flags for CO_2 and u* < 6 (Vellinga et al., 2013).

[1] Webb, Pearman,& Leuning (Webb et al., 1980).
[2] Based on aircraft attitude and solar position.
[3] Integral Turbulence Characteristics

14. Line 365 & 528

To my knowledge Metzger et al. (2022) [https://doi.org/10.5194/amt-14-6929-2021] were the first to use combined airborne and tower EC heat and water flux data in ML applications to optimize flight track placement. Clarify how the current work extends this to in-situ CO2 fluxes.

Noted. We changed on line 385: "Under similar circumstances, when the tower dataset lacks spatial and temporal coverage, we believe the inclusion of airborne data can improve the model substantially. This was also demonstrated by Metzger et al. (2021), who showed that airborne measurements in combination with pre-field simulation experiments doubled the potential of a surface–atmosphere study. However, in the current study, as the tower network had been seriously extended resulting in much higher spatial coverage, this was not directly visible in an improved R^2. We examined the airborne data's added value in three ways."

On line 528, we specified that we are the first (to our knowledge) to combine these data sources for CO_2 fluxes: "We merged CO_2 flux data from both airborne and tower measurements, and, to our knowledge, this study is the first to use this combination of CO_2 data as input for a machine learning model." We also added a reference to Metzger et al. (2013) in the introduction as a study where airborne and ground-based measurements are combined in a ML framework.

References

Aben, R. C. H., Craats, D. V. D., Boonman, J., Peeters, S. H., Vriend, B., Boonman, C. C. F., Velde, Y. V. D., Erkens, G., & Berg, M. V. D. (2024). Using automated transparent chambers to quantify CO2 emissions and potential emission reduction by water infiltration systems in drained coastal peatlands in the Netherlands. *EGUsphere*, *403*. https://doi.org/10.5194/egusphere-2024-403

Boonman, J., Hefting, M. M., Huissteden, C. J. V., Berg, M. V. D., Huissteden, J. V., Erkens, G., Melman, R., & Velde, Y. V. D. (2022). Cutting peatland CO2 emissions with water management practices. *Biogeosciences*, *19*, 5707–5727. https://doi.org/10.5194/bg-19-5707-2022

Buzacott, A. J., Kruijt, B., Bataille, L., van Giersbergen, Q., Heuts, T. S., Fritz, C., Nouta, R., Erkens, G., Boonman, J., van den Berg, M., van Huissteden, J., & van der Velde, Y. (2024). Drivers and annual totals of methane emissions from dutch peatlands. *Global Change Biology*, *30*. https://doi.org/10.1111/gcb.17590

Casso-Torralba, P., Vilà-Guerau de Arellano, J., Bosveld, F., Soler, M. R., Vermeulen, A., Werner, C., & Moors, E. (2008). Diurnal and vertical variability of the sensible heat and carbon dioxide budgets in the atmospheric surface layer. *Journal of Geophysical Research: Atmospheres*, *113*(D12).

de Arellano, J. V.-G., Gioli, B., Miglietta, F., Jonker, H. J., Baltink, H. K., Hutjes, R. W., & Holtslag, A. A. (2004). Entrainment process of carbon dioxide in the atmospheric boundary layer. *Journal of Geophysical Research: Atmospheres*, *109*(D18).

Evans, C. D., Peacock, M., Baird, A. J., Artz, R. R., Burden, A., Callaghan, N., Chapman, P. J., Cooper, H. M., Coyle, M., Craig, E., Cumming, A., Dixon, S., Gauci, V., Grayson, R. P., Helfter, C., Heppell, C. M., Holden, J., Jones, D. L., Kaduk, J., … Morrison, R. (2021). Overriding water table control on managed peatland greenhouse gas emissions. *Nature*, *593*(7860), 548–552. https://doi.org/10.1038/s41586-021-03523-1

Foken, T., Göockede, M., Mauder, M., Mahrt, L., Amiro, B., & Munger, W. (2004). Post-field data quality control. In *Handbook of micrometeorology: A guide for surface flux measurement and analysis* (pp. 181–208). Springer.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., … Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the fluxcom approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Koch, J., Elsgaard, L., Greve, M. H., Gyldenkærne, S., Hermansen, C., Levin, G., Wu, S., & Stisen, S. (2023). Water-table-driven greenhouse gas emission estimates guide peatland restoration at national scale. *Biogeosciences*, *20*, 2387–2403. https://doi.org/10.5194/bg-20-2387-2023

Kruijt, B., Buzacott, A., van Giersbergen, Q., Bataille, L., Biermann, J., Berghuis, H., Heuts, T., Jans, W., Nouta, R., & van Huissteden Hefting F Hoogland, J. M. (2023). Co2 emissions from peatlands in the netherlands: Drivers of variability in eddy covariance fluxes.

McDermitt, D., Burba, G., Xu, L., Anderson, T., Komissarov, A., Riensche, B., Schedlbauer, J., Starr, G., Zona, D., Oechel, W., et al. (2011). A new low-power, open-path instrument for measuring methane flux by eddy covariance. *Applied Physics B*, *102*, 391–405.

Meesters, A., Tolk, L., Peters, W., Hutjes, R., Vellinga, O., Elbers, J., Vermeulen, A., Van der Laan, S., Neubert, R., Meijer, H., et al. (2012). Inverse carbon dioxide flux estimates for the netherlands. *Journal of Geophysical Research: Atmospheres*, *117*(D20).

Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón Y Widemann, B., Neidl, F., Schäfer, K., Wieneke, S., Zheng, X. H., Schmid, H. P., & Foken, T. (2013). Spatially explicit regionalization of airborne flux measurements using environmental response functions. *Biogeosciences*, *10*(4), 2193–2217. https://doi.org/10.5194/bg-10-2193-2013

Metzger, S., Durden, D., Paleri, S., Sühring, M., Butterworth, B. J., Florian, C., Mauder, M., Plummer, D. M., Wanner, L., Xu, K., & Desai, A. R. (2021). Novel approach to observing system simulation experiments improves information gain of surface–atmosphere field measurements. *Atmospheric Measurement Techniques*, *14*(11), 6929–6954. https://doi.org/10.5194/amt-14-6929-2021

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., & Yakir, D. (2006). Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583.

Vellinga, O. S., Dobosy, R. J., Dumas, E. J., Gioli, B., Elbers, J. A., & Hutjes, R. W. (2013). Calibration and Quality Assurance of Flux Observations from a Small Research Aircraft. *J. Atmos. Ocean. Technol.*, *30*(2), 161–181. https://doi.org/10.1175/JTECH-D-11-00138.1

Vellinga, O. S., Gioli, B., Elbers, J. A., Holtslag, A. A. M., Kabat, P., & Hutjes, R. W. A. (2010). Regional carbon dioxide and energy fluxes from airborne observations using flight-path segmentation based on landscape characteristics. *Biogeosciences*, *7*, 1307–1321. www.biogeosciences.net/7/1307/2010/

Vickers, D., & Mahrt, L. (1997). Quality control and flux sampling problems for tower and aircraft data. *Journal of atmospheric and oceanic technology*, *14*(3), 512–526.

Wageningen Environmental Research. (2024). Bodemkaart van nederland [Accessed: 2025-04-11]. https://www.bodemdata.nl

Webb, E. K., Pearman, G. I., & Leuning, R. (1980). Correction of flux measurements for density effects due to heat and water vapour transfer. *Quarterly Journal of the Royal Meteorological Society*, *106*(447), 85–100.