Answer to referee comment 1 for “Quantifying agricultural N_2O and CH_4 emissions in the Netherlands using an airborne eddy covariance system”

We thank reviewer 1 for the evaluation of our manuscript, especially for pinpointing sections with possibility to improve clarity. Below we list our answers to the comments and corresponding revisions in the manuscript. Referee comments are written in normal font, our answers in italic font.

The present study "Quantifying agricultural N2O and CH4 emissions in the Netherlands using an airborne eddy covariance system" describes in detail a measurement system and airborne eddy covariance method for quantifying CH4 and N2O fluxes over an agricultural region. The study is generally clearly presented and extremely detailed in methodology. Novel findings of N2O emissions from agriculture and comparison to inventories are also presented. The authors thoroughly consider uncertainties and potential biases in the measurements. I recommend publication following minor revisions.

Line 32: I believe the current IPCC recommendation for the GWP100 of biogenic CH4 is 27.

Thank you for this correction, 27 is the most recent value IPCC assessed for GWP100 of biogenic CH4. We changed the corresponding sentence in the manuscript.

General comments/questions, largely for the purposes of improved clarity:

1: I assume that the choice of 90 s was at least partially made to capture all of the eddy scales based on ogives, analysis of the cospectral power, and/or integral timescale. I don't believe an explanation of this was explicitly given in the manuscript. It would be helpful to see in the text a description of what factors went into to choosing the 90 s windows.

The choice of an appropriate window length for flux calculation is an important aspect of our work, and not straight forward, especially not for airborne applications. Hence, we derived and evaluated the window length: Assumption “zero” was based on typical ground-based EC flux calculation intervals. Those are typically 30 min in length. We have translated those 30 min timescales into a length-scale by multiplying it with typical horizontal wind speeds (3-6 m/s), yielding 5-11 km. With our average aircraft speed of 62 m/s, we need to fly 90 – 180 s to capture the same (5-11 km) length scale. Hence, 90 s was our first minimum window length to start with. Then, as presented in Section 3.1, we used cospectral analysis and the ogive method to test whether our 90 s window length is sufficient or not. As demonstrated by the ogives for all cases, even for those with long flux legs (large window sizes) we don’t see significant flux contributions at the low-frequency end of the spectrum (frequencies below 0.014Hz). Therefore, we can conclude, that our window length is long enough to capture all relevant flux carrying eddies.

We adapted corresponding paragraphs in Sections 2.1.2 and 3.1 to improve the description of our approach of choosing the 90 s window length.

2: I found it generally confusing what is meant by leg, vs Flight leg, vs flux segment. It would be useful if clear definitions were explicitly given and/or more consistency in the language were used. e.g. Do these terms always refer to the flight leg, or sometimes to the 90 s intervals as it seemed?

Thank you for this hint. To avoid confusion by inconsistent terminology, we changed our terms to "flight path", when ever talking about a certain part of the whole flight pattern (e.g. between two turns). When referring to a certain distance, for which a single flux value is calculated, we introduced the term "flux leg". Flux leg replaces the former terms of leg, flux segment and further. The temporal pendant to flux leg is the size of the moving window.

3. The authors mention that spatial homogeneity is required for eddy covariance, but the flux variations over a leg seem to indicate non-homogeneity. I would assume the condition of homogeneity is only necessary over the 90 s windows use for the flux calculations. Does the overlapping windows further loosen this condition? Some discussion of this in the text would be useful.

We indeed found a significant spatial variability of fluxes across the target area (Hotspots vs. areas with low fluxes), hence heterogeneity in the emission strength. The eddy covariance method formally requires homogeneity of the underlying flux field, but as you point out, only across the window for which the flux is calculated. This implies stationary turbulence, a constant footprint area and a spatially uniform surface flux within the flux leg. In practice, for airborne fluxes this condition is rarely met for entire target areas, since flight tracks inevitably cross heterogeneous landscapes. The moving-window approach allows us to resolve spatial variability of fluxes along the flight track, but it also implies a relaxation of the EC assumption of surface homogeneity. Each window provides a flux estimate under the assumption of local homogeneity within the corresponding footprint, while the overall heterogeneity of the landscape is captured by the variability across consecutive windows. Variability between consecutive windows can then be interpreted as spatial heterogeneity in the flux field, and is not a violation of the EC assumption of homogeneity.

We added sort of this discussion of flux heterogeneity and its relation to the moving window approach partly to the description of the moving window technique in the manuscript Section 2.1.2:

"Short flux legs allow for higher spatial resolution and are more likely to be located within a homogeneous source area, but potentially miss flux contributions from large eddies (see Section 2.1.3) and are associated with higher flux uncertainties because of smaller sample sizes (see Section 3.2). Overly long flux legs are prone to violations of source homogeneity and do not offer fine spatial resolution."

And partly to section 4.1, where we discuss spatial variability of fluxes:

"Repeated flux legs over the same ground scene, but flown during different times of the flights, and even at different times of the day (morning flight 21a and afternoon flight 21b), yield similar emission rates. This reproducibility indicates spatial coherence and temporal consistency of the flux signals, hence stationarity. We observe smooth transitions in fluxes of single flux legs between regions of low fluxes and regions of high fluxes and since consecutive flux legs have a large spatial overlap with corresponding similar footprint, local homogeneity within the flux legs can be assumed. Thus, the variability of N_2O fluxes across the target area, from close to zero in some parts of the area up to emissions of around 1 μg m^{-2} s^{-1} (in the central part of the patterns) demonstrates the ability of our airborne EC setup to detect and hence, spatially resolve small-scale (i.e. 1–10 km) emission hotspots within a relatively homogeneous landscape."

4. Are the LODs calculated for each 90 s segment (i.e. N represents the number of observations per 90 s flux interval)? If so, are these simply averaged over the flight? If, rather, N is the entire leg, wouldn't this calculation of LOD underestimate the ability of the instrumentation to

distinguished spatially-resolved fluxes? In general I think more clarification is needed to contextualize the LODs reported.

Yes, LODs are calculated for each window (90 s segment). We use the moving window approach to achieve high spatial resolution by inspecting fluxes of consecutive windows, hence we need to define the LOD of each single window to assess whether the corresponding flux value can reliably be differentiated from zero.

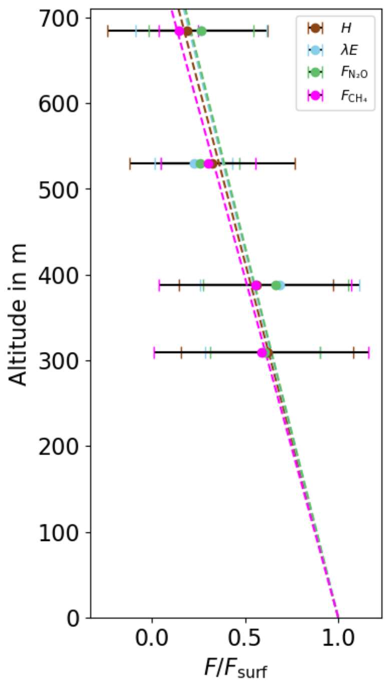
We clarified in the manuscript that LODs were calculated for each single window and complemented the manuscript by the following paragraph, with the aim to provide a better insight to the meaning of flux LODs for the reader: Beyond using LODs to distinguish between pure random noise and true flux, they can be used to compare different EC systems based on their smallest detectable flux. However, this comparison is somewhat limited, since the LOD depends on instrumental noise, turbulence conditions, and window length, hence varies across different flux legs, as turbulence intensity and scalar variance can change, which directly propagate into the LOD calculation. It is therefore not only system specific, but also dependent on turbulence and the true flux field. Thus, single flux leg LODs are no constants but should be interpreted as context-dependent thresholds of confidence. By averaging LODs over multiple flux legs and across different flights, we can achieve a better comparability to other airborne EC systems.

Specifically, we added the following paragraph:

"Beyond using LODs to distinguish between pure random noise and true flux, they can be used to compare different EC systems based on their smallest detectable flux. However, this comparison is somewhat limited, since the LOD depends on instrumental noise, turbulence conditions, and window length, hence varies across different flux legs, as turbulence intensity and scalar variance can change, which directly propagate into the LOD calculation. Furthermore, the LOD reduces by averaging repeated overflights of the same flux leg (Langford et al., 2015). LODs are therefore not only system specific, but also dependent on turbulence and the true flux field. Thus, single flux leg LODs are no constants but should be interpreted as context-dependent thresholds of confidence. By averaging LODs over multiple flux legs and across different flights, we can achieve a better comparability to other airborne EC systems. Our setup achieves a relatively low averaged LOD for CH_4 fluxes of 0.14 $\mu g\ m^{-2}\ s^{-1}$. This is comparable to the range reported by Wiekenkamp et al. (2025) (0.1−0.14 $\mu g\ m^{-2}\ s^{-1}$), and notably better than the average LOD of 0.66 $\mu g\ m^{-2}\ s^{-1}$ reported by Pasternak (2023). To our knowledge, the only other airborne EC N_2O flux study, estimated an LOD of 0.1 $\mu g\ m^{-2}\ s^{-1}$ (Wilkerson et al., 2019). Our system achieves an lower value of 0.037 $\mu g\ m^{-2}\ s^{-1}$."

5. It would be helpful to have a figure on the flux divergence calculation in the Appendix.

We put the following figure into the Appendix and used it to explain how we assessed the vertical flux divergence correction, by adding the text paragraph beyond.



"Figure A1 shows measured averaged fluxes of H, λE, CH_4 and N_2O, normalized by their extrapolated surface flux values, for the different altitude levels probed during the vertical flux divergence experiment. Similar slopes of normalized fluxes of all scalars indicate a gradual decrease of measured fluxes from the surface upwards. z_i was inferred to be 890 m amsl. on this day (21 June 2023) by vertical profiles."

6. Authors mention recent studies utilizing the continuous wavelet transform method, which is often thought of as better for obtaining higher spatial resolution. Is there a reason that the authors used the moving window method instead?

Continuous wavelet transform (CWT) is sometimes applied in airborne flux measurements and offers advantages in cases of strong spatial heterogeneity or non-stationary conditions with e.g. intermittent turbulence. Its main benefit is the high resolution in both frequency and time domain. We used classical EC partly because we compensate for the superiority of CWT by using appropriate moving windows: They allow for high spatial resolution, limit the assumption of stationarity to only the short windows, and with cospectral analysis, we demonstrate that we don't suffer from low-frequency loss.

On the other hand, despite the advantages, the use of CWT requires careful consideration of several factors, including the low-frequency cutoff for cospectrum integration. This low-frequency cut-off (using Fourier transformation) is also part of the discussion in our work, and by default, with CWT one integrates contributions from all scales, conceivably non-turbulent large-scale contributions, which must not to be interpreted as instantaneous surface flux (Li et al., 2023). Secondly, the choice of an appropriate mother wavelet (e.g., Morlet or Mexican hat) is not always clear and can lead to different quality of the flux calculation under different conditions (Mi et al., 2005, Schaller et al., 2017). Third, the validity of flux estimates near the edges of flux legs (cone of influence) has to be considered. The overall aim of this study is to demonstrate that agricultural greenhouse gas fluxes can be reliably quantified with the new airborne system. For N_2O fluxes, this is a new approach to our knowledge. Therefore, we have chosen classical eddy covariance as an appropriate first step, given its long-standing establishment and acceptance in the flux measurement community, before applying the more advanced CWT.

We have added a short statement (to not confuse the reader with methodological details), why we have chosen classical EC in the paragraph where we mentioned CWT:

"Continuous wavelet transform analysis holds the potential to overcome the trade-off between spatial resolution and uncertainty (Vaughan et al., 2021; Wolfe et al., 2018; Metzger et al., 2013), but its application is outside the scope of this study, as its implementation requires careful methodological choices and interpretation, while the classical EC approach provides a well-established and reliable framework for this first demonstration of our system."

References

Mi, X., Ren, H., Ouyang, Z., Wei, W., and Ma, K.: The use of the Mexican Hat and the Morlet wavelets for detection of ecological patterns, Plant Ecology, 179, 1–19, https://doi.org/10.1007/s11258-004-5089-4 , 2005.

Schaller, C., Göckede, M., and Foken, T.: Flux calculation of short turbulent events – comparison of three methods, Atmospheric Measurement Techniques, 10, 869–880, https://doi.org/10.5194/amt-10-869-2017 , 2017.

Li, Y., Wu, Y., Tang, J., Zhu, P., Gao, Z., and Yang, Y.: Quantitative Evaluation of Wavelet Analysis Method for Turbulent Flux Calculation of Non-Stationary Series, Geophysical Research Letters, 50, e2022GL101 591, https://doi.org/10.1029/2022GL101591 , 2023.