Response to Access Reviews:

Reviewer #1:

We would like to thank reviewer #1 for the positive evaluation, the identification of the small glitches in the text and thus implemented all the suggest corrections. Our responses are given in italic font.

Technical corrections requested by Reviewer #1:

Typing corrections:
Line 30: year "2014" is written as "214" - *DONE*
Line 47: "as well" should be "as well as" - *DONE*
Line 48: "… during all" seems an incomplete sentence – *DONE (we added "years")*
Line 132: "ecosystem" should be "ecosystems" - *DONE*
Line 202: Fuchs et al reference states the years as "2081" – presumably this should be "2018" (also use of square brackets?) - *DONE*
Line 345: Should "bon" be "box"? - *DONE*
Table 1: Max data availability for 2010, 2011, 2013 & 2014 should be 17520 not 17521
Table 4: Please differentiate Parcel A and B for Fertilizer and Harvest (both currently labelled as Parcel A) - *DONE*
Figure 4 caption: please remove "(a)" immediately after "Figure 4:" - *DONE*

Figure and Table Clarifications:
All applicable tables and figures: please indicate sign convention either in the caption or on the figure (e.g. Figure 2, that positive values represent and export) - *DONE*
Figure 3: Please indicate what the grey lines represent (presumably uncertainty or standard deviation?) - *DONE*
Figure 4a: Please state in the caption what the grey lines represent (half-hours fluxes?) - *DONE*

Reviewer #2:

We would like to thank reviewer #2 for the overall positive evaluation and for providing feedback on the points that the reviewer encourages to be addressed. For this access review, we implemented all technical corrections right away and lay out a way forward on how we intend to address the currently raised points following the end of the full review period in a fully revised version of the manuscript. Our responses are given in italic font.

The manuscript "Memory effects on greenhouse gas emissions (CO2, N2O and CH4) following grassland restoration?" by Merbold et al. is a well written long term study of GHGs from a grazed grassland system in Switzerland. The team have used a mixture of measurement methods over a 5 year period to get a very good picture of a full GHG budget for the field. This is a very valuable study as such long term observations are rare and it answers some questions that are not well studied.
I found the manuscript interesting to read, and it was written to a very high standard and I do believe that it should be published after some amendments.
I do have some comments that I feel should be addressed by the authors that I believe would improve the quality and usefulness of the study for others. Although these comments

are numerous and not entirely simple to address, if the authors can amend their study to incorporate them I feel the work would benefit greatly.

Thank you for the positive evaluation and we suggest ways forward point by point below.

1. A large assumption made by the study is that the eddy covariance measurements are entirely truthful of the conditions in the field. It has been observed in the past that long-term carbon budgets derived from eddy covariance can be biased due to assumptions made by the method. Often negative carbon fluxes are reported in similar systems, however, when investigating deep soil cores there was found to be no significant difference in C content of the soil (see Jones et al., doi:10.5194/bg-14-2069-2017 for one such study). The manuscript does not provide evidence of the C stock in the soil beyond the Eddy C measurements to back up the evidence which would have made it a much more significant study. This does not invalidate the study by any means, but without clarification of potential uncertainties, it increases the danger that the study provides "concrete" evidence of mitigation methods (i.e. grazing animals is a carbon sink) that has been used recently by advocates of the meat industry to justify the long-term environmental aspects of livestock farming. I would advise a short message of discussion to highlight that there is room for error in the measurements and that soil carbon was not measured to validate the measurements. Alternatively, if the soil measurements are there, please include them.

Indeed, we agree that soil inventories should be linked to EC fluxes more often, particularly since EC measurements are often seen as entirely truthful. We are confident that the EC method is a valuable and powerful tool to investigate C and N fluxes at ecosystem scale – not necessarily the exact entire field as suggested in the literature (Hill et al. 2016 https://doi.org/10.1111/gcb.13547) – yet, it allows to derive a general view on whether a system is likely to gain/loose carbon/nitrogen. We further agree that continuous flux measurements and thus budgets should be validated with other independent methods, ie a soil inventory. Yet, determining changes in soil C/N is similarly not trivial and takes considerable time as suggested by ie. Schrumpf et al. 2011 https://doi.org/10.5194/bg-8-1193-2011. Additional, within this specific project we were not able to carry out a resampling of the soils while further advocating for this in follow-up projects. Certainly, we will aim at implementing a thorough uncertainty assessment of the numbers presented in the revised version of the manuscript. Multiple approaches to estimate the uncertainty in EC flux measurements as well as in gap-filling methods are available (ie Post et al. 2015 https://doi.org/10.5194/bg-12-1205-2015, Vitale et al 2019 https://doi.org/10.1007/s00477-019-01664-4, Hollinger and Richardson 2005 https://doi.org/10.1093/treephys/25.7.873, Nicolini et al. 2018, https://doi.org/10.1016/j.agrformet.2017.09.025)

2. I do not agree with the way that the N2O flux data has been handled in the study. N2O fluxes measured using chambers almost always follow a log-normal distribution in space, so any data analysis must take this into account when handling means and uncertainties. A simple arithmetic mean with associated uncertainty (not sure what the error bars on Fig 3 and 4 represent?) will not be an adequate way to represent

this data (although commonly used wrongly in previous studies). This will result in a skewing of the data and large overestimates in minimum confidence intervals and underestimations of maximum confidence intervals. An example is when uncertainties of N2O cross the negative threshold when no observations of flux dip below zero. This is not a satisfactory way to present the data. I recommend using a more sophisticated analysis technique and showing 95% confidence intervals where possible for a thorough comparison of the measurement techniques.

We thank the reviewer for the critical assessment. Our approach followed the method used by Hoertnagl et al. 2018 https://doi.org/10.1111/gcb.14079. We already added here – as technical correction - the information on the grey lines (standard deviation). Furthermore, and following the full review process we anticipate to take the log-normal distribution into account.

3. L303: Due to the log-normal distribution of N2O emissions measured using chambers, most measurements will be very close to zero and ppb differences in gas samples will hover around detection limits of the analysis instrument. In such cases, the R2 value of the fits will be very low for many, but the regression between points will still be valid (effectively an average of the instrument noise with a slope near zero). By cutting data with R2 lower than 0.8 I assume that a very large number of small fluxes are removed from the dataset. If this is the case I would recommend a threshold on this QC method, or a more detailed explanation of what impact this had on the data in the text if this is not the case (as I read it, the method would likely contribute to a large bias is flux estimates).

We implemented thorough QC criteria concerning the N2O flux calculations. All the details have been in detail provided in Imer et al. 2013 https://doi.org/10.5194/bg-10-5931-2013 stating the r2 threshold of 0.8. Overall, the low fluxes being part of our observations were not being the limit of detection and have thus been included in this study. In the revised manuscript we will further clarify out QA/QC approaches in order to avoid misinterpretation.

4. Uncertainties in cumulative emissions are not presented which makes it difficult to compare with other studies or what impact gap-filling and weather may have had on the study. This should be easily manageable for CO2 for which models exist, and probably for CH4 using simple gap-filling as it was found to be approximate zero throughout the study. I understand that there is no definitive way to gap-fill N2O, however a running median is not a statistically defensible way to "model" data. As a result no uncertainty will be calculated from this method. If the authors want to estimate uncertainties in cumulative N2O fluxes, they will have to develop a more sophisticated approach to gap-filling.

We agree with the reviewer and will implement such uncertainty estimates, in particular for CO2 and CH4, in the revised version of the manuscript following the full review process. The Running median approach was chosen, following Hoertnagl et al. 2018 (see above) as this seemed at the time being the best possible way to fill N2O flux data gaps. In terms of

uncertainty, we intend to present additional gap-filling methods for N2O in the revised manuscript.

5. I feel a nitrogen budget without NH3, NOx and N2 is not very useful. Combined, these gases will likely contribute approx. 50% of nitrogen losses from the system. Perhaps a better way to confer N losses is to calculate the emission factors of the fertiliser applications, as that is a more generally used term for such activities in literature and is a better description of the presented results in the study.

We are in full agreement with the reviewer that other N compounds build a large part of the N budget. In order to proceed we suggest two points: (1) We will further clarify that we are only showing a partial N budget caused by the fact that we do not have data available for NH3, NOx and N2 losses, and (2) to incorporate emission factors based on the losses observed via N2O and the fertilizer inputs in the revised manuscript.

6. Is there a way to estimate the N content of the fodder/grass on the field before tillage to assess the emissions from the herbage being tilled into the soil?

We have thought about this too when preparing the manuscript and must confess that we had not taken such measurements. However, to our current knowledge the additional N being incorporated into soil during tillage should be small caused by relatively low vegetation cover/height at this time of the year. For the revised manuscript, we will provide a better estimate based on existing literature.

7. Does the carbon budget take into account vehicle use? Is it insignificant or does tractor diesel have a role to play?

The currently presented budget does not include C emissions from vehicle use for two reasons: (1) the hours farm vehicles are being used on this field are very limited over the course of the year given the small size of the fields (negligible). The negligibility of these emissions was further underlined (2) by a MSc thesis that investigated full farmgate budgets in the years prior this study. We will add an additional sentence concerning this point in the revised version of the manuscript.

8. L225: Can you explain what you mean by an internal reference cell in the instrument for the QCLAS? To my knowledge, these cells are used to find absorption lines on the spectra and not for calibration as they leak over time. The QCLAS system typically does not require calibration as it operates on the principles that the absorption follows Hitran quantum mechanics laws.

Thank you for this comment and this seems to be a misunderstanding of what we have written. We stated that the infrared gas analyser was calibrated regularly, while we also wrote that the QCLAS was fitted against an internal reference cell. In order to create better clarity we changed this sentence as follows: "The QCLAS did not need calibration due to its

operating principles, and an internal reference cell (mini-QCL manual, Aerodyne Research Inc., Billerica, MA, USA) eased finding the absorption spectra after each restart of the analyzer."

Some minor corrections
L283: I think there is a bit of wording here that is confusing. Flushing the chamber with the syringe isn't technically correct. I think it would be better to say that the syringe was used to pump the chamber to circulate the air to avoid the concentration gradients? - *DONE*
L471: here the order of the sentences makes it sound like CH4 contributed to 70% of the budget. Please re-order. – *DONE, we added "*the contribution of CO_2*"*
L606: Change highlight to highlights - *DONE*