We would like to thank the referee for his comments and respond as follows:

“An intercomparison of eddy covariance and chamber methods is like comparing apples and oranges”: We agree with the referee, chamber and eddy covariance methods are very different approaches to measure N2O fluxes. However, both methods are used to measure N2O fluxes in field experiments with the aim of calculating annual budgets, which will contribute to the improvement of national N2O emission factors. We therefore think that there is a need of a comparison of both methods and that a comparison is justified. As discussed in the paper, both methods have their limitations
and flux calculations always have a degree of uncertainty attached. This uncertainty, however, can be constrained by comparing fluxes calculated by these two very different methods. We appreciate that there are differences between the two approaches; hence we suggest settling for a comparison of oranges and mandarins...

‘Differences between N2O fluxes calculated from chamber measurements and EC over the same measurement period were never significant’: The sentence is referring to the comparisons of median and mean fluxes over time from EC comparison points and chamber measurements (page 12, line 5). The variability of the two methods is therefore not included, but only the variability of the measurements over time. We understand that the sentence in the abstract might be misleading as the connection to the temporal variability is not given. We therefore decided to remove the sentence.

Chamber setup in the field: The chambers were removed every two weeks to reduce the chamber effect on the vegetation and soil and allow free grazing. Chambers were re-positioned at least 24 hours before measurement, to avoid the influence of the soil disturbance on N2O production. The grass inside the chambers was always accessible to the animals for grazing, apart from the 1 hour period during which chambers were closed for the N2O measurements. Grazing maintained a canopy height that was always lower than the chamber height (20cm) and therefore chambers were operational throughout the inter-comparison periods. Due to the accessibility of the chambers to grazing and the changing of the position every fortnight, the influence of litter as an energy source for denitrifiers was the same for the area measured with chambers as for the area influencing EC measurements. We will include this additional information in the manuscript.

Discrepancy in time series and scatter plots: The time series (Figure 2) and scatter plots (Figure 3) are complementary but not contradictory. In the time series for May 2007 the timing of the peak is not different between EC and chamber measurement; the increase, maximum and decrease of flux magnitudes occur on the same day for each method. Only the magnitude of fluxes is different, which is something that is generally
not highlighted in a scatter plot. The same applies for July 2007, if only comparison points are considered. On the 19th of May 2007, no chamber measurements were carried out; they could have measured similar high fluxes as EC measurements. On the 16th of July 2007, no data for EC were available; they could have measured similar high fluxes as the chamber. We thank the referee for the comment and have tested the use of the second lowest chamber instead of the average of all four chambers. This did not improve the correspondence in the scatter plot apart from March 2007 (all r²=0.42, 2nd lowest r²=0.48). In spite of this, your comment that the key influence acting against a better correlation is the spatial variability of the chambers is correct. A better correlation could only be achieved by including many more chambers in such an intercomparison.

Negative fluxes measured with the EC setup: The authors thank the referee for this comment. As we write in the discussion section (4.2, page 17, line 22-24) we do not believe that the high negative fluxes we measured with the EC method reflect true exchange fluxes, especially isolated spikes. The reason we did not discard those negative peaks so far, was that we did not want to bias the dataset by only discarding negative values and leaving in positive ones that were possibly affected by the same methodological artefact. However, we agree with the referee’s comment and we decided to apply more restrictive filtering protocols to ensure that all instantaneous fluxes presented in the paper are reliable. In particular, the spiky fluxes will be investigated with further spectral analysis and the newly filtered fluxes will be presented in the revised manuscript. The lag time showed a fairly stable behaviour through the different field campaigns and it has otherwise been used as a rejection criterion for identifying suspicious fluxes. The authors have applied a stationarity criterion (see Affre, C., Lopez, A., Carrara, A., Druilhet, A., Fontan, J., 2000. The analysis of energy and ozone flux data from the LANDES experiment. Atmospheric Environment 34, 803–821.), which is conceptually very close to the cumulative spectrum method the referee refers to. We thank the referee for pointing this out, as it wasn’t mentioned in the original method section. We will include this additional information in the manuscript. Although we agree that
sudden negative fluxes are possibly due to instability of the EC setup as a whole, we believe that the detection limit should be based only on the systematic errors of the system, not on random errors.