Dear editor and reviewer,

We would like to thank you again for taking the time to look through our manuscript. I apologise that my initial reply was not satisfactory in dealing with your concerns regarding the two points that the reviewer reiterates. I hope that the changes I have made now full address these issues.

In response to the reviewers comments,

In my previous review I pointed out that there was an interesting pattern in the yield data and suggested that it might be worth to explore other statistical approaches. I understand that separating the data set is not the right way but the analyses need to take into account the variability between fields. For example, it might be worth to work with response ratios, a way to make data comparable between fields, or do a mixed effect model including "field" as a random effect.

I understand that the reviewer would like to see more clearly how the fertiliser works differently on the different fields. I have added a paragraph with the suggested yield response ratios to further highlight the differences between the two fields used in the study.

Text added:

" In terms of yield response to the fertilisers, large differences were observed between the two adjacent experimental fields, even though historical management practices were largely similar. In the Engineer's field plots (2016), the response to the fertiliser was muted, with relatively large variation between the plots. Yield response (and standard deviation) of the plots (treated minus control) was largest for the nitram treatments at 19 (± 10) %, while the urea and inhibitor treated urea had little impact on crop yield, with only a 2.0 (± 23) % and 0.7 (± 26) % larger harvest when compared to the control plots, respectively. In the Upper Joiner field (2017), the yield response was much higher at 150 (± 144), 113 (± 69) and 136 (± 107) % for the nitram, urea and treated urea treatments."

Regarding Figure 4, I was asking if there are no error bars and suggesting that there might be an outlier in their data set, making it impossible to compare 2016 Event 2 to the others. 500 mg N/kg for ammonium just seems a lot. Doing a back-of-the-envelope calculation: The fertilizer application of 70 kg N/ha equals roughly 30 mg N/kg, assuming a bulk density of 0.7 g/cm3 and 30 cm soil depth. This means that the ammonium concentration at that one time point is approximately 15-fold higher than the fertilizer added. Anyway, it would have been easier to assess this if there would be some sort of measure of the variability of the replicated plots displayed in the figure. The authors argue that it is exceptionally difficult to display meaningful error bars. In my opinion, showing the mean and the standard deviation of the four replicates is meaningful. Figure 3 shows each plot, which helps to assess the variability of the replicated plots. Figure 4 also needs to show somehow the variability of the replicated plots.

Regarding the high ammonium in the soil, this is due to individual measurements, which we reanalysed and got the same results. Five of the samples measured NH_4 concentrations above 1000 mg/kg of the 1000+ samples during the project. But it would not be correct to regard these as "outliers as they follow a log-normal distribution (Figure 1 below) showing decreasing probability towards high values, but a consistent curve from about 0 to values over 1000. In previous experiments we have also seen measurements on this magnitude, following a log-normal scale to very high concentrations, and it is difficult to pinpoint exactly what causes them. As well as natural mineralisation processes and mineralisation of previous manure residue in the plots, the actions of earth worms and urination of wild animals (potentially deer, rabbits and resting birds breaking into the plots), occasionally the core just happens to measure a bit of soil that has a few fertiliser pellets land within a few cm of each other, so it is highly concentrated at that particular location. NH_4 will not travel through the soil as nitrate would as it is fairly immobile in the soil, so it's more likely that the NH_4 would remain concentrated in the first few cm where the crops and microbes will be unable to assimilate it under the chemical conditions. All of these factors make it very difficult to show what is happening in the soil with the data collected, as it only takes one high value in a set of plots to skew everything for any particular plot/fertilisation. We could potentially do something like cut off all values above 250 mg/kg as "outliers" to help deal with the data, but as with our studies on N_2O data handling have shown, dealing with log-normal data appropriately is important.

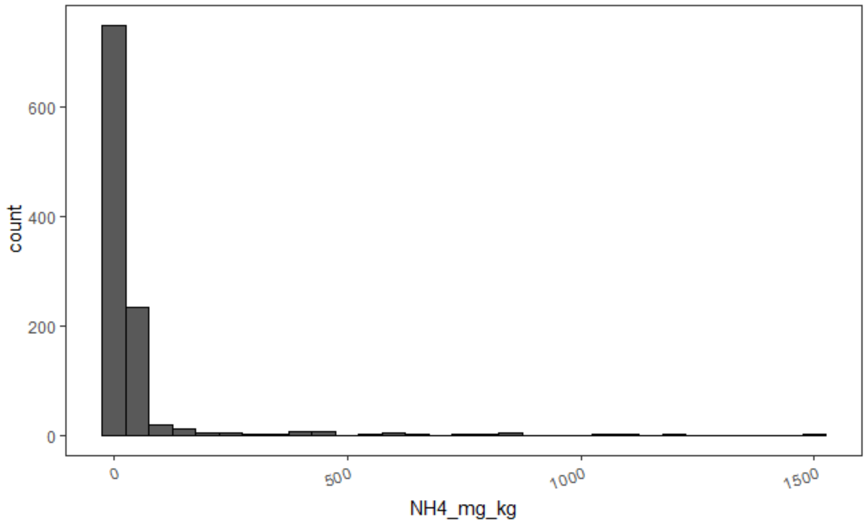

Figure 1 Histogram of all NH4 measurements

I include the following text in the discussion:

"The reason for such high individual concentration measurements of available Nr measured sporadically throughout the experiment are unclear, and may arise as a result of a cluster of fertiliser pellets dissolving in close proximity, or due to outside influences such as urine patches from rabbits or other wild animals. Similar spatial variability in available nitrogen is observed at the field scale in local studies (Cowan et al., 2015; Cowan et al., 2017) and may be due to residues of animal excreta present in the soils."

My reservation with the error bars asked for is that, as the data is skewed (but still real), standard deviations no longer apply as the data is not normalised, so we end up with some strange error bars that cross well over into negative values even though no negative measurements are observed (See figures below). These are not really representative of the data either. The range of the errors also hide what is happening to the mean over time as the scales are so large. Plotting the all of the measurements individually would stretch the y-axis further, making the plot not very meaningful. Essentially, we need to apply even more Bayesian statistics to handle the data, but there is no model published that we can use to predict what is happening, so we are stuck with a difficult data set.

One major difficulty is presenting the data in a way that a non-specialist reader can observe the measurements, without having to think about log-space transformations. Our first attempt was to keep it simple and use medians, which to some extent show changes without overcomplicating the data. In order to satisfy the above request I attempted a table, but it's not possible to fit all the data in for all the events/dates without a very large and impractical table. So I have edited the manuscript to include the conventional way of handling data as requested, plotting the mean and std devs of the data (in ribbon format as below).

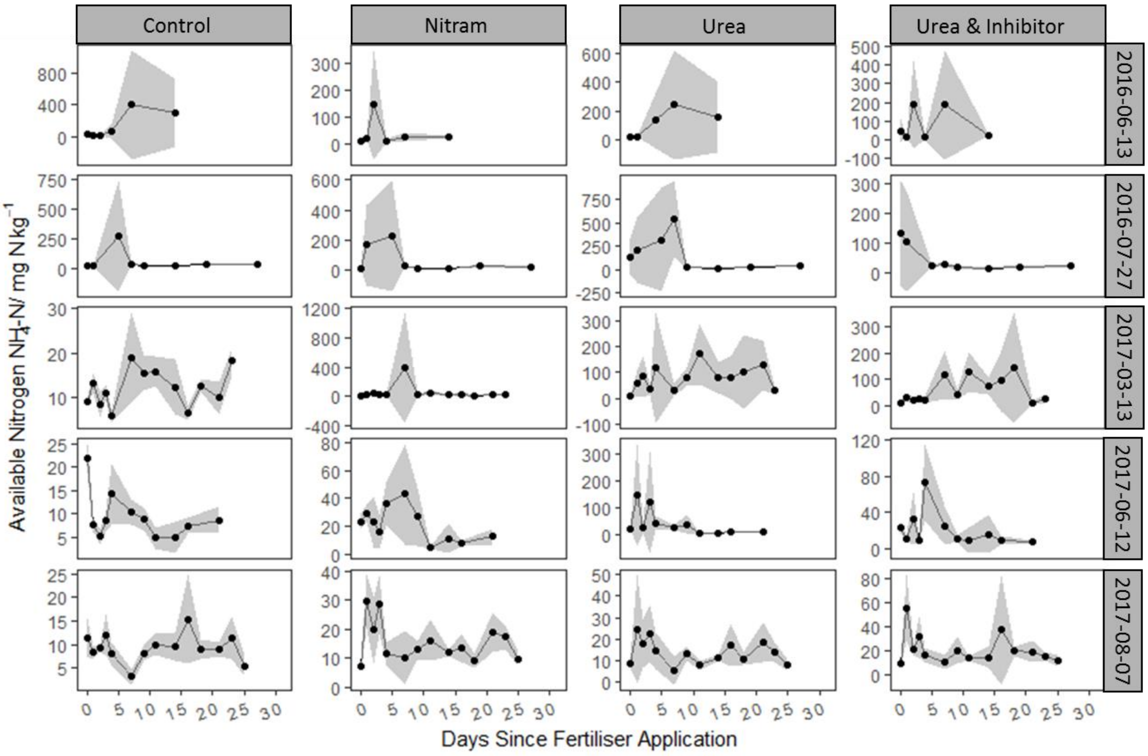

Figure 4 Mean ammonium concentrations from soil samples (n= 4) measured in tandem with N_2O chamber measurements after fertilisation events. Standard deviation is included (grey ribbon).

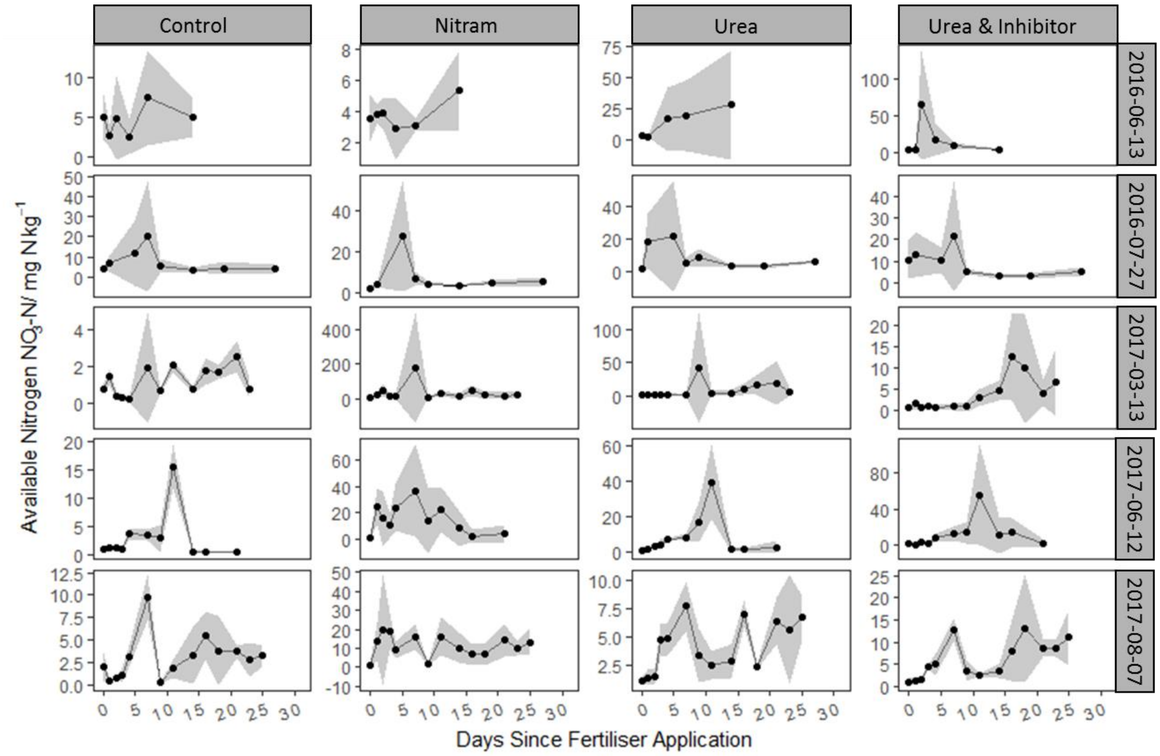

Figure 5 Mean nitrate concentrations from soil samples (n= 4) measured in tandem with N_2O chamber measurements after fertilisation events. Standard deviation is included (grey ribbon).

We thank the reviewers and the editor again for their input and comments which we feel have definitely improved the manuscript, and we hope that these final corrections are adequate to see the submission through the review process.