We thank the reviewer for his/her comments. In the following please find our point-by-point response to these comments, marked with an "R"

The authors estimate N_2O emissions for the US Corn Belt using a top-down approach. The authors have made a valiant effort, but I think the paper still has a long way to go. Below I have included a number of broad, high level suggestions and line-specific suggestions.

High-level suggestions: - In the introduction, I would be sure to emphasize the new or novel scientific questions that you want to tackle in the article. These questions should be different from existing N_2O studies. In other words, I would focus on the scientific gaps that these existing studies have not been able to answer.

- Make sure the motivations and scientific questions in the introduction connect with the questions that you answer in the discussion/conclusions. For example, there is a lot of text in the intro about Eulerian versus Lagrangian models and the new perspective that an Eulerian model would provide. However, it's not clear from the discussion and conclusion sections what we learned from this Eulerian model that we could not have learned from a Lagrangian model.

R: We have added the following texts to emphasize the contribution of the present study:

'To our best knowledge, this study appears to be the first one that uses WRF-Chem to do inverse analysis for N_2O, analyzes the influences of monitoring height on the inverse analysis results, and illustrates the spatial characteristics of the influences of the Corn Belt on the atmospheric N_2O concentration. Deployment of measurements made at multiple heights in inverse analysis can reduce the uncertainty in modeling transport and diffusion in the atmospheric boundary layer and therefore should provide better estimates of the actual emission than if only one measurement height is used'.

The reviewer's comment regarding STILT refinement will be discussed in the following responses.

I felt that the methods were convoluted, particularly given the simplicity of the topdown approach. For example, the paper appears to use both preset scaling factors and an empirically derived scaling factor (unless I misunderstood the text). This combination of both pre-set and empirical scaling factors feels too complicated, given that the top-down strategy in the paper is ultimately very simple. See the specific suggestions below for more detail on this point.

R: The 'scaled simulations' with the preset scaling factors (3, 6, 12, and 25) are used to find the parameter a in inversion Equation (1): $M_C - 1 = a\ (M_F - 1)$. In Fig. 6b, we compared observation with the modeling results using multiplier 25, because 25 is close to the constrained scaler of 28.1.

In the revised manuscript, we have added the following explanation,

'The scaled simulations with multipliers of 3, 6, 12, and 25 are used to find parameter *a* in Equation (1)'. (lines 15-16, page 5)

To clearly describe the inverse method, we have modified Fig. 3 slightly, and add an example in the caption to explain how the flux multiplier was constrained.

'The hollow dots and dot lines in sub-figure (e) show how the emission flux multiplier M_F (the dot on x axis) is determined via the concentration multiplier M_C (the dot on y axis) for October 1-20'.

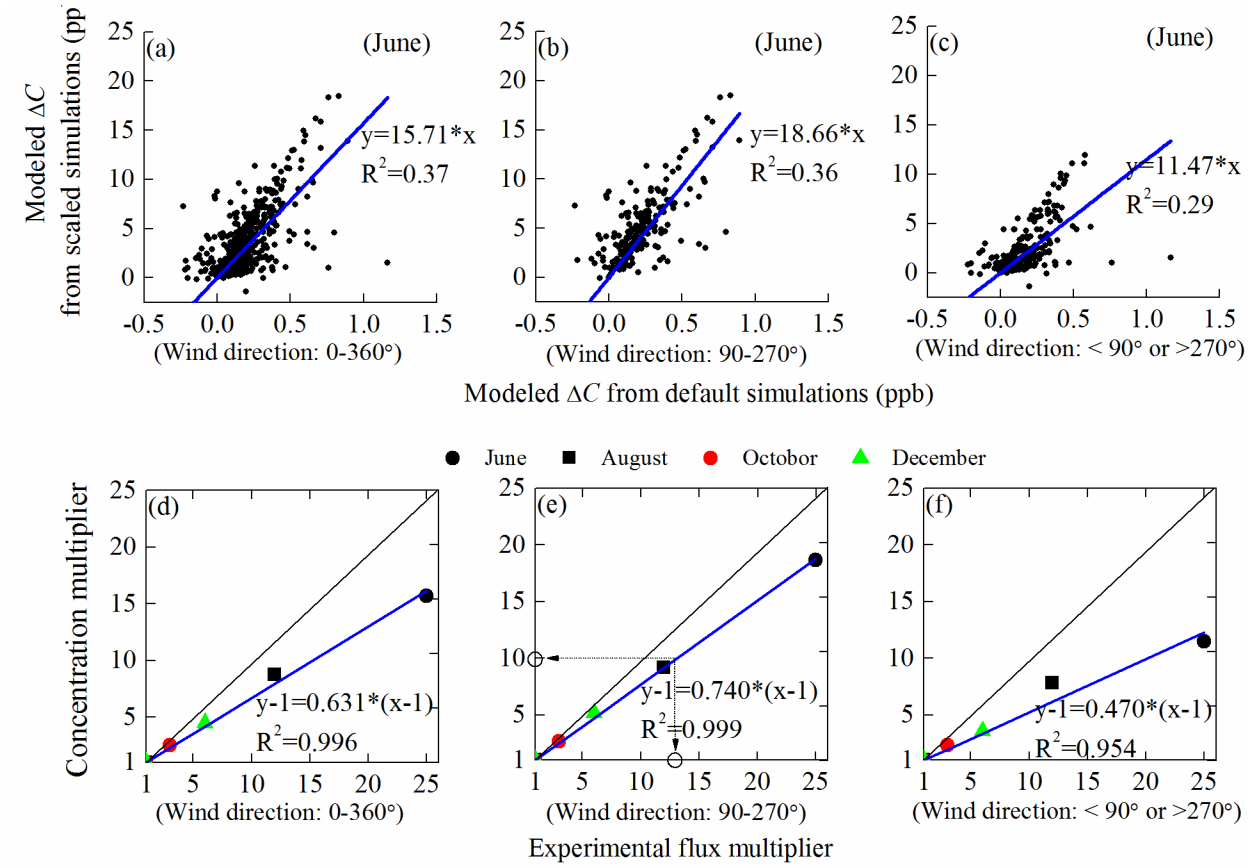

Figure 3. Relationships for different model runs between concentration multiplier and experimental flux multiplier. The modeled N_2O mixing ratio enhancement ΔC was obtained from default and scaled simulations for 185 m at the KCMP tower. The scale simulation shown in panels a – c uses a multiplier of 25.0. The regression slope in panels a – c is represented by the black circle in panels d – f. The hollow dots and dot lines in sub-figure (e) show how the emission flux multiplier M_F (the dot on x axis) is determined via the concentration multiplier M_C (the dot on y axis) for October 1-20.

We also provided a numerical example in the online supplement to explain the steps involved in the inverse calculation.

- Try to use technically specific phrases whenever possible and try to avoid broad generalizations. For example, the phrase "model accuracy" appears multiple times in the text, but it's not clear what aspect of the model framework that phrase refers to.

R: We have changed the section name 'Model accuracy' to 'Mixing height', and changed 'model accuracy' where it appears for the first time to 'accuracy of the modeled atmospheric N_2O transport' in the revised manuscript. (lines 5-6, page 6)

Specific suggestions:

Pg 1, Line 16: "evaluated" might be a better word choice here than "validated"

R: It has been changed as suggested. (line 16, page 1)

Pg 1, Line 20: You compare your flux numbers against EDGAR in the abstract. You could also consider putting these numbers in the context of existing top-down studies of US N_2O emissions (of which there are several). Do you get similar or different numbers? In the latter case, what factors might explain these differences?

R: Yes, we have compared our inversion results with other inversion studies conducted for the central U. S. or Corn Belt, and the differences have also been explained, as shown in section '4.2 Comparison with other emissions estimates'. (from line 13, page 10 to line 13, page 11)

Pg 1, Line 33: "emissions are". I think it's more standard to use the plural ("emissions") and not the singular ("emission"). This suggestion applies to multiple instances of the word throughout the text.

R: This has been corrected throughout the revised manuscript. Thank you.

Pg 1, Line 36: I would specify whether these methods are used by your study or by existing studies. I.e., the topic sentence doesn't make it clear whether are you are about to do a literature review or whether you are about to describe the current study. Consider using active voice here instead of passive voice.

R: ‘in previous studies’ has been added to the end of this sentence in the revised manuscript. (line 2, page 2)

Pg 2, line 19 "is usually simulated": I think you could be a bit more specific here. E.g., "All of the top-down studies described here used STILT ....". Alternately, "All but one of the top-down studies ....".

R: Following the reviewer’s suggestion, we have changed this sentence to ‘All top-down inverse modeling studies described here used …’. (line 22, page 2)

Pg 2, line 21: "computational"

R: Corrected. (line 24, page 2)

Pg 2, line 23: Why is it important to quantify the "spatial characteristics of atmospheric N_2O mixing ratios"? I would argue that a model like STILT could be used for this task by running a large number of receptors or footprints. However, I don’t think there has ever been a need to do so in the context of in situ greenhouse gas observations.

R: It is possible to run STILT for a large number of receptors, but to do so for 44×34 and 189×98 grid cells (in the outer and inner domains respectively) and 40 vertical layers would be very time consuming.

In response to this comment, we have added the following explanatory text:

“Currently, there are only a few stations monitoring atmospheric N_2O concentration in the Corn Belt. Detailed information on the spatial distribution of N_2O can help experimentalists position their observational sites strategically. It also reveals the spatial extent of the influences of local emissions on the atmosphere”. (lines 24-26, page 8)

Pg 2, line 25: Again, why is it important to explore "the relationship between the spatial characteristics of surface emissions and the atmospheric N_2O mixing ratio at the regional scale".

R: Please see our response above.

Pg 2, line 26: Why would an Eulerian model be any better at quantifying turbulence than a model like WRF-STILT? If you keep this motivation in the introduction, I would expand the discussion to answer this question.

R: We believe that the two modeling strategies are complementary to each other. We have revised the text to:

"It is recognized that some of the parameterizations in STILT, such as the turbulent velocity variance and the Lagrangian timescale, need refinement to improve model performance (Pillai et al. 2012). On the other hand, Eulerian models cannot distinguish the contribution of a specific source to the atmospheric contribution. Overall, both Lagrangian (e.g., STILT) and Eulerian (e.g., WRF-Chem) models have their advantages and disadvantages in inverse analysis, and comparing their results obtained from the same region can inform refinement efforts on these models". (lines 29-34, page 2)

Since we did not run the STILT model, we were not able to directly compare the two modeling strategies. In the discussion section (Section 4.2), we did compare our inversion results with those obtained by STILT modeling.

Pg 2, line 34: What do you mean by "force agreement"? In what way are you forcing this agreement? Instead of using a phrase like this one, I would instead indicate whether your simple inverse model is a scaling factor inversion, a grid-scale inversion, a geostatistical inversion, etc.

R: This sentence has been rewritten as:

"estimate the actual emissions as some multiple of the agricultural N_2O emissions in EDGAR42 whereby the multiplier was obtained from these relationships and the observed atmospheric N_2O on a tall tower in Minnesota".

Pg 2, lines 36-40: This study sounds very similar to Chen et al. I would try to delineate as clearly as possible what lingering science questions you want to answer in this study. For example, you mention analyzing the "influences of monitoring height on the inverse analysis results." You could elaborate on this point and explain how that analysis would benefit the community or answer important science questions.

R: We have added the following texts to emphasize the contribution of the present study:

'To our best knowledge, this study appears to be the first one that uses WRF-Chem to do inverse analysis for N_2O, analyzes the influences of monitoring height on the inverse analysis results, and illustrates the spatial characteristics of the influences of the Corn Belt on the atmospheric N_2O concentration. Deployment of measurements made at multiple heights in inverse analysis

can reduce the uncertainty in modeling transport and diffusion in the atmospheric boundary layer and therefore should provide better estimates of the actual emission than if only one measurement height is used'. (lines 10-14, page 3)

Pg 3, line 20: Why use Niwot Ridge as the background in this study? In theory, one could use any number of different sites in the NOAA network as the background. Alternately, one could also use Arlyn Andrews' empirical boundary curtain product.

R: We did not try Arlyn Andrews' empirical boundary curtain product. Instead, the initial and boundary conditions for the N_2O mixing ratio for each modeling period in the present study were obtained from Model for Ozone and Related Chemical Tracers (MOZART) version 4. We select Niwot Ridge as the observation background site for four reasons: (a) it is located at the upwind side of the Corn Belt; (b) it is basically outside of the zone of influence of the Corn Belt emissions (Fig. 6); (c) observations at Niwot Ridge are continuous in time whereas observations at other sites are made only once or twice a day; (d) we want to be consistent with former studies, i.e., Griffis et al. (2013) and Chen et al. (2016).

In the revised manuscript, we added the following text:

'Another advantage of using NWR as opposed other NOAA monitoring sites is that the observation at NWR is continuous in time whereas measurements at other sites are made only once or twice per day. Griffis et al. (2013) and Chen et al. (2016) also used the observation at NWR as the background concentration'. (lines 22-24 in page 5)

To analyze the uncertainty caused by representation of the background mixing ratio, we have done additional inverse analysis using AMT (Argyle, located in Maine) as the background site. The two background sites show slightly different concentrations (Figure R1). The absolute change of the constrained multiplier is 1.4 – 1.9, 1.9 – 3.3, 1.1 – 1.8, and 0 – 0.1 for 1^{st} – 20^{th} in June, August, October, and December, respectively (Table R1). The inverse uncertainty caused by background selection is larger in August and October than that in June and December.

These new results are included in the online supplement (Figure S5 and Table S4).

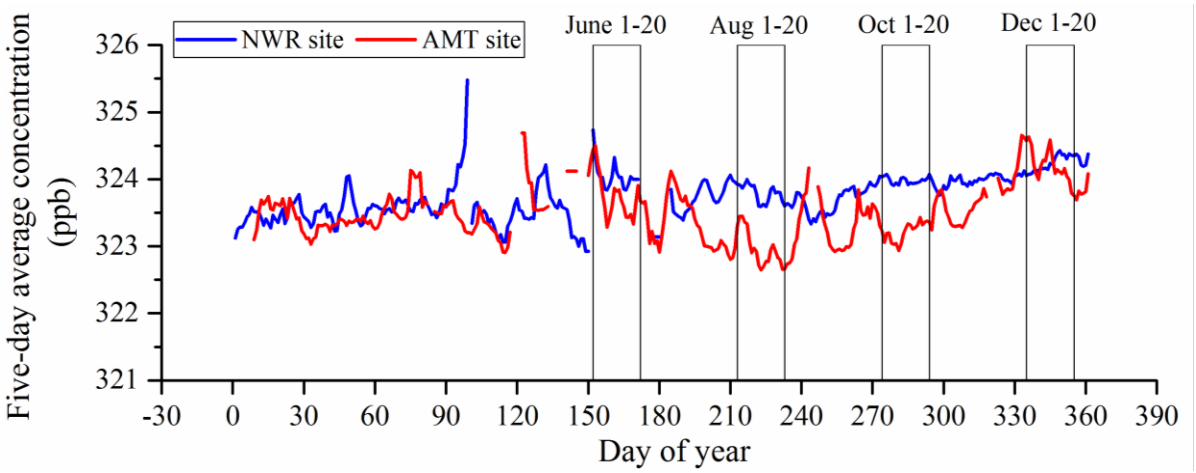

Figure R1. Observed N_2O mixing ratio at the Niwot Ridge and "Argyle, Maine" (AMT) sites.

Table R1. Experimental and optimized flux multiplier M_F **using AMT as background site**. Values in brackets are constrained agricultural emission flux in units of nmol m^{-2} s^{-1}.

Time	June 1 – 20	August 1 – 20	October 1 – 20	December 1 – 20
Experimental	0, 1, 25	0, 1, 12	0, 1, 3	0, 1, 6
Optimized [a]	20.4 (3.12)	11.2 (1.71)	4.5 (0.68)	3.0 (0.47)
Optimized [b]	24.3 (3.71)	14.0 (2.15)	5.2 (0.80)	3.7 (0.57)
Optimized [c]	30.0 (4.59)	16.3 (2.49)	6.5 (0.99)	4.4 (0.67)

Notes: a, b, c: using observation data at heights of 32, 100, and 185 m, respectively.

Table 2 in the present study. Experimental and optimized flux multiplier M_F. Values in brackets are constrained agricultural emission flux in units of nmol m^{-2} s^{-1}.

Time	June 1 – 20	August 1 – 20	October 1 – 20	December 1 – 20
Experimental	0, 1, 25	0, 1, 12	0, 1, 3	0, 1, 6
Optimized [a]	19.0 (2.91)	9.3 (1.43)	3.4 (0.52)	3.0 (0.47)
Optimized [b]	22.5 (3.44)	11.6 (1.77)	3.8 (0.59)	3.6 (0.55)
Optimized [c]	28.1 (4.29)	13.0 (1.99)	4.7 (0.72)	4.3 (0.66)

Notes: a, b, c: using observation data at heights of 32, 100, and 185 m, respectively.
We have discussed the uncertainty introduced by the selection of the background mixing ratio in the newly added section '4.3 Other sources of uncertainty'.

Pg 3, paragraph starting with line 25: In many STILT studies, the meteorological model (e.g., WRF) has a different spatial resolution than the STILT footprint. This paragraph mixes and matches the resolution of the meteorology model and the resolution of the footprints. I would clarify which of these two cases you are referring to.

R: we use WRF-Chem to simulate the N_2O transport (as a passive tracer): the meteorological field modeled by WRF and the concentration field modeled by the chemistry module of WRF-

Chem share the same grid cells and have the same resolution. In this respect, our model is different from the WRF+STILT method.

In the revised manuscript, we have added the following sentence: 'The meteorology module and chemistry module use the same mesh generation'. (line 10, page 4 in the revised manuscript)

Pg 4, line 25: Why use these pre-set multipliers?

R: please refer the response to the general comments above.

Pg 4, paragraph starting with line 28: I'm confused by the methodology here. Why use set multiplier values (as described in the previous paragraph) and then fit a coefficient ("a") to a simulation that has already been scaled by some pre-set multiplier? Instead, I would fit modeled concentrations (using some inventory) to the observations with a simple regression. I'm not sure what additional leverage one gets by using the more complicated setup described here.

R: We have rewritten the text here to improve clarity:

"The preset multipliers represent our first guess values. The actual multiplier values are constrained by the concentration observations at the KCMP tower. First, we define…

We used the observed enhancement, ΔC, defined as the actual concentration observed at the KCMP tall tower minus a background concentration and adjusted for a small spatial gradient in the modeled N_2O mixing ratio between KCMP and the background concentration site from the background simulation, to constrain the flux multiplier (and the surface emission flux). A numerical example is given in the Supplementary Information on how this is done."

In the Online Supplement, we now provide a numerical example to show the steps involved in this calculation.

We agree with the reviewer that an alternative approach is to constrain the flux with a regression relationship between the modeled and the observed concentrations. A slight advantage of the present approach is that it reveals that there is not a 1:1 correspondence between concentration enhancement and the emission enhancement: As shown in Fig. 3e, when the emission strength is increased by 3, 6, 12, and 25 times, the modeled concentration enhancement is increased by 2.7, 5.1, 9.1, and 18.7 times, respectively. We believe that this information is useful to the modeling community.

Pg 5, line 5-7: I disagree with the statement here. One could generate footprints or sensitivities either using an Eulerian model adjoint or using a model like STILT. That procedure would circumvent the need to run a model in iterative fashion, like what the authors describe here. Also, this statement clashes with the introduction. In the introduction, the authors argue that Eulerian models provide a vital perspective that Lagrangian models cannot provide. But here, the authors point out a big shortfall of their Eulerian approach that would not be true of a Lagrangian approach. I would tone down the language in the introduction accordingly. I would also be more specific in lines 5-7. For example, if WRF-Chem does not have a readily-available adjoint, that argument would be more compelling.

R: Although not deployed here, we agree that model adjoint is a useful tool for analyzing the sensitivity of a model output variable to several input variables spatially and simultaneously. We also agree that STILT could generate footprints. We have pointed out that 'An advantage of using STILT to conduct the inverse modeling for N_2O is that it needs much less computational resources than full three-dimensional Eulerian models'. (lines 23-25, page 2)

We have also toned down the language in the introduction (see our response above).

General comment: The material in section 2.3 also relates to inverse analysis. I would merge sections 2.3 and 2.4.

R: these two sections have been merged in the revised manuscript.

Page 5, beginning of section 3: You may want to give you reader a road map/outline of this section, telling your reader what information to expect.

R: the following outline has been added in the revised manuscript,

'In this section, we first compare the modeled mixing height with those derived from other data products to evaluate the accuracy of the modeled N_2O transport in the atmosphere. We then discuss the relationship between modeled concentration enhancement and the flux enhancement; these results are used to establish Equation (1). The constrained emission flux values are given next. After that, we compare the modeled mixing ratio with the hourly observations at the KCMP tower and the modeled mixing ratio spatial distribution with those observed at multiple NOAA flask sites. Finally, we present the spatial distribution of the modeled atmospheric N_2O mixing ratio in the Corn Belt.'. (from line 37, page 5 to line 4, page 6)

Page 6, line 25: This statement feels too broad or general to me. What kind of model performance are you referring to here? Transport model performance? The performance of your estimated emissions?

R: this sentence has been changed to 'Accurate assessment of the model performance in reproducing the atmospheric N_2O mixing ratio is difficult at hourly intervals …'. (lines 5-6, page 8)

Page 7, paragraph beginning with line 4: I think that analysis here is complicated by the experimental setup that has the pre-set scaling factors. I.e., the low bias is the result of the particular scaling factors that you chose. Again, I would use a simple regression in sections 2.3 and 2.4 and focus on the results that lead to direct scientific conclusions.

R: Since we did not run a simulation using a multiplier of 28.1, we chose to show the results from the scale simulation with the preset multiplier of 25.0 which is very close to the optimized multiplier of 28.1.

We have modified the text slightly to

"Two reasons may explain this low bias. The first minor reason is that the results shown in Fig. 6 were from the scaled simulation with an experimental flux multiplier of 25.0, and this multiplier is slightly smaller than the optimized flux multiplier M_F value of 28.1 determined in post-simulation analysis. Second,…"

Page 7, line 14: What kind of "model accuracy" are you referring to here? Atmospheric transport, accuracy of the emissions, spatial/temporal aggregation errors, or something else?

R: this sentence has been changed to 'A key factor in inverse analysis is the accuracy of the modeled atmospheric N_2O transport and turbulent mixing' in the revised manuscript. (line 6, page 6)

Section 4.1: This section is mostly devoted to validating your atmospheric modeling setup. It feels like this material is more well-suited to a supplement, not to the beginning of the discussion section. I.e., this section does not tell us much about N_2O but rather explains why you think your approach is a valid one.

R: We have moved the 'Model accuracy' contents in original manuscript to the '3 Results' section in the revised manuscript, based on the comments from the first reviewer. Specifically,

we have moved the PBL height analysis to section '3.1 Mixing height' and the spatial comparison to section '3.5 Spatial variations of modeled N_2O concentration'.

Page 8, line 25: I wouldn't use nighttime measurements at either 32m or 100m. You could potentially use nighttime measurements at 185m; that inlet height may lie above the nocturnal boundary layer. But I think even using nighttime measurements at 185m is really challenging/ill-advised.

R: We believe (and Reviewer 1 agrees) that sub-day variations carry important information about the surface source strength and atmospheric mixing. Kim et al. (2013) also used both the daytime and nighttime observations at the height of 100 m on the KCMP tower to do inverse analysis for CO. Because nighttime measurements, especially those made near the surface, are subject to stronger influences of the sources closer to the tower, the optimized fluxes may be biased heavily by those sources. This concern is not too serious in the present study, for two reasons. First, the land use classes (and presumably emission patterns) near the tower are very similar to those in the whole Corn Belt (Griffis et al. 2013). Second, the EDGAR emission strength is not sensitive to the distance from the tower, as shown in Figure R2 below. In Figure R2, the agricultural emission strength and the fertilization rate are plotted as functions of distance from the KCMP tower in the south wind sector (90° – 270°). Within the radius of about 500 km from the tower, the emission strength and the fertilizer application rate are approximately constant with distance (Figure R2), noting that 500 km is the distance from the KCMP tower to the south boundary of the Corn Belt.

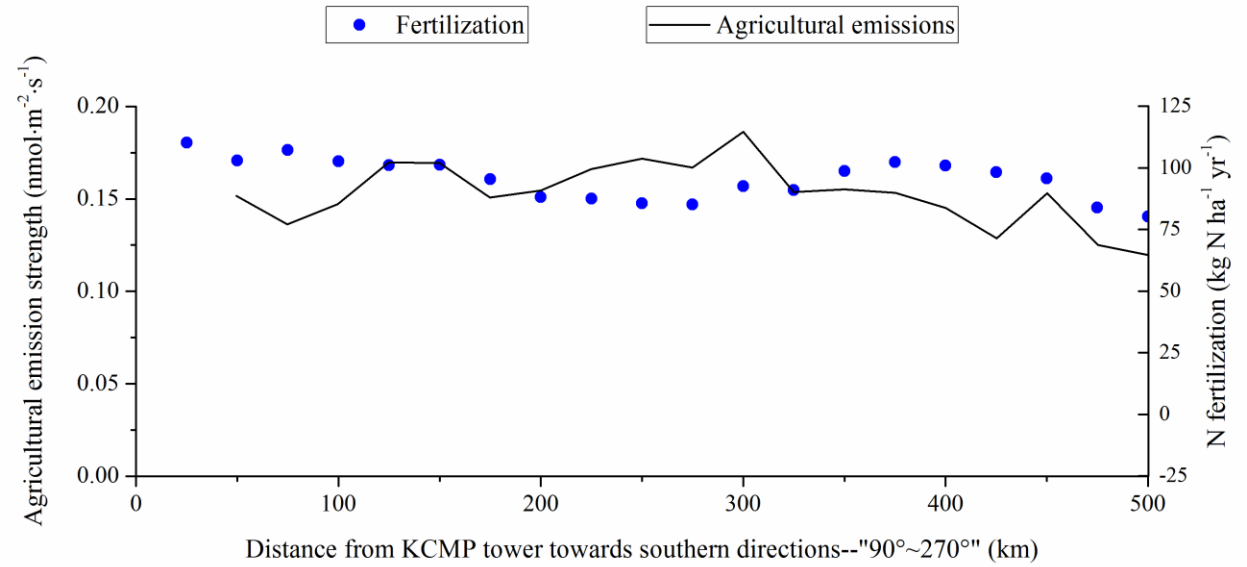

Figure R2. Agricultural N_2O emission strength versus N fertilization from the KCMP tower toward south (90° – 270°).

Page 9, lines 21-25: I disagree with this statement, at least in part. The argument here might be valid for Kort et al. 2008. They used aircraft data from across the US and did estimate a single scaling factor. Miller et a. 2012, by contrast, used a gridscale inversion (not just a single scaling factor), and they used tall tower data from the central US agricultural belt. Also, I don't think it's guaranteed that Kort et al.'s scaling factor would have been larger had they used data specific to the corn belt. I think it's a possibility but not a definitive explanation as presented here.

R: Thank you. We have changed 'a single emission scaling' to 'the emission scaling' in the revised manuscript. (line 23 in page 10)

In addition, we have also changed 'the resulting emission flux would have been larger' to 'the resulting emission flux would probably be larger'. (line 25-26 in page 10)