We thank the Reviewer for taking efforts to go through our manuscript and providing his/her comments. The point by-point responses (bold text) to the comments (normal text) are given below and corresponding changes made are highlighted in red color in the revised manuscript.

General: The study investigates simulated ozone over South Asia, using several simulation scenarios, composed of different inventories and chemical mechanisms. The simulation results were evaluated using data from an in-situ monitoring network. Among the findings of the study is that simulated daytime ozone maximum differ significantly between different emission scenarios, by as high as -22%, in contrast to the 24h mean values, which are more consistent. The results are not surprising, especially on local scale, given that measured ozone is primarily photo-chemically formed. However, a major issue here is that the authors use different temporal emissions (2010 for HTAP, 2006 for INTEX-B) form different emission inventories and are trying to validate the model simulations of 2013 (using reanalysis ECMWF product) with measurements from completely different temporal period (e.g, 2004 or before, and 2009-2013), except for 4 stations. The authors should clarify the significances of these results in this context, especially in this very active developing region? Impacts from biomass-burning emissions are not adequately discussed. The authors proclaim similar results between different emissions scenarios despite the different temporal periods. However, these claimed similarities should be only a warning of some compensating effects that cancel the interesting differences caused by the emissions annual trends and variability. The study sounds scientifically interesting and well written, but still need more consistent analysis and casual discussions on the driving factors of the differences between these scenarios.

We thank the reviewer for careful evaluation of our manuscript and constructive comments. Considering the lack of high-resolution measurements and bottom-up emission inventories for different years, the current understanding of the spatio-temporal distribution of surface ozone (Kumar et al., 2012b; Ojha et al., 2016; Ansari et al., 2016; Girach et al., 2017) and its impacts on crop yield (Ghude et al., 2014) and human health (Ghude et al., 2016) are based on WRF-Chem simulations driven by one of the inventories coupled to RADM2 or MOZART chemistry, or by averaging several simulations. However, there does not exist a comprehensive information about how different are the modeled ozone levels among different emissions and chemistry options being used in the aforementioned studies.

We agree that the emissions over this region are changing, however time dependent bottom up inventories are not available for all years, and the inventory of a different year is commonly used (Kumar et al., 2012b; Kumar et al., 2015; Ghude et al., 2016; Ojha et al., 2016). Therefore, it is important to inter-compare ozone simulated using different inventories. The numerical experiments performed for a common year with varying employed inventories can provide generalized but very important first hand information about how much variablility exists in simulated ozone if one inventory is used as compared to other. Similar to the effects of inventories, choosing a different chemical mechanism also has considerable effects on simulated ozone. Therefore, keeping every other input fixed we vary the chemical mechanisms to report the differences that change of chemical mechanism causes. Thus we do not believe that the limitations raised by reviewer, which may be valid within themselves, dilute our scientific conclusions in any way.

An additional challenge in simulating the ozone pollution in this part of the world is the lack of in situ high resolution data in time and space to validate model output. Previous evaluation of a 2008 model run (driven by emissions representative of 2006) had to rely on datasets older by 10

years or more (Kumar et al., 2012b). We tried to minimize the temporal differences in model and observations by conducting new observations in rapidly developing Delhi and another station in Pune and using the data for recent years (2009-2013) compared to previous evaluation efforts at several stations. Finally the effect of different meteorological year on estimated biases is studied and discussed (Page: 9, Lines: 319-326). We believe and mention in the revised manuscript that a compilation effort such as ours will provide a scientific basis to stress on making continuous observations over a network of stations, and making it available through projects such as TOAR (http://toar-data.fz-juelich.de/). This is discussed in the revised manuscript (Page:14, Lines: 541-543).

The point by-point responses to the specific comments are given below in bold.

Comment 1: Page 1, lines 32-33. The conclusion that the SEAC4RS-RADM2 scenario preforms better than the others does not sound novel scientific information. I think that it is important here that the authors shed some light on why this specific scenario works better than the others.

Response 1: Model evaluation and inter-comparison studies such as these serve as a reference for subsequent usage of model to address scientific questions. The intercomparison experiments presented in this paper show that the current understanding of the ozone budget and implications for human health and crop yields have large uncertainties over India. Additionally, the information that SEAC4RS-RADM2 simulations are in better agreement with observations have implications for future studies to minimize the aforementioned uncertainties. Previous studies analysing crop loss and mortality due to ozone exposure have not explicitly considered the comprehensive and detailed evaluation performed in this study. Thus the aim of our study is to fill this gap of information on model evaluation which is to be considered by the scientific community to study and control crop losses and pre-mature mortalities due to ozone exposure.

Comment 2: Page 3, lines 103. The authors mentioned high pollution loading and biomass burning as reasons for the intense ozone photochemical formation during the pre-monsoon period. It would be also very interesting if the authors could investigate how biomass burning emissions and transport affect ozone photochemical formation in the study's domain.

Response 2: The effects of biomass burning on ozone over Indian region have been studied by Jena et al. (2014) reporting O_3 enhancement by 4-10 ppb (25-50%) in the Eastern region including Burma, 1-3 ppb (10-25%) in Central India and 1-7 ppb (4-10%) in the Indo-Gangetic region. Further, O_3 enhancement was found to be about 2-6 ppb (8-20%) over the Bay of Bengal in March, which was attributed to the transport from the Eastern region. As suggested by the reviewer, this is now discussed in the revised version of the manuscript (Page: 3 ; Lines: 106-109).

Comment 3: Page 4, lines 139-141: Could the authors elaborate on the difference between the two aerosol modules used, the (MADE/ SORGAM) vs GOCART, and how this would affect their results?

Response 3: We reiterate that the aerosol-radiation feedback is kept off in this study, to investigate the effects specific to emissions of O_3 precursors (Page: 6; Line: 203-204), therefore a different aerosol module would not impact the results significantly. A similar procedure had been utilised previously to compare emissions inventories for modelled ozone over Southeast Asia (Amnuaylojaroen et al., 2014).

Comment 4: Page 4, lines 142-145: Also, how the different photolysis schemes Fast-J and F-TUV may affect the results? Could the authors employ the same aerosol and photolysis scheme for each scenario (using different emissions and chemical mechanism), so that casual factors for the differences can be determined?

Response 4: While comparing the simulations with different emissions (HTAP-RADM2, INTEX-RADM2 and S4RS-RADM2), the aerosol mechanism and the photolysis scheme are kept same, so differences between the three runs can be attributed to the differences in emissions.

Because of the way the two mechanisms RADM2 and MOZART are implemented into WRF-Chem, they use different photolysis schemes: RADM2 uses the Madronich TUV or Fast-J scheme, and MOZART uses the "Fast" TUV (Madronich F-TUV) scheme, which is based on the same physics as the Madronich TUV scheme, but designed to run faster. The differences between the two Madronich photolysis schemes is further described in the supplementary material to Mar et al. 2016.

In the present study although RADM2 uses the Fast-J photolysis scheme, a sensitivity simulation with Madronich TUV scheme revealed similar surface ozone mixing ratios and chemical tendencies at various model levels with small differences (<5%) over most of Indian region (not shown). So our results would be similar if we use Madronich TUV scheme instead of Fast-J scheme with RADM2. Further, Mar et al. (2016) used Madronich TUV scheme with RADM2 and Madronich F-TUV scheme with MOZART chemical mechanism and reported that the two different Madronich photolysis schemes had only a small contribution to the differences in the predicted ozone by two chemical mechanisms. The major difference between two chemical mechanisms was due to differences in inorganic reaction rates (Mar et al, 2016). Hence we conclude that in our study too, the differences over Indian region are primarily due to choice of the chemical mechanisms irrespective of photolysis scheme used. Moreover, as the aerosol radiation feedback is turned off hence the observed differences are mainly result of differing gas phase chemistry. This is discussed and clarified in the revised version (Page: 11; Lines: 394-405).

Comment 5: Page 4, line 152: What is the effect of using year 2010 HTAP emissions as opposed to experimental observation date and model reanalysis of 2013? How this may affect their conclusions?

Response 5: As explained in the manuscript, to evaluate this effect we conduct an additional simulation for 2010, and find only small differences in the estimated model biases (±3 ppbv in 3 years) and our results are not affected significantly (see supplementary Fig. S4; Page: 9, Lines: 319-326).

Comment 6: Page 5, line 160: What is the effect of using year 2006 INTEX-B emissions as opposed to experimental observation date and model reanalysis of 2013? How the authors account for using emissions from different years?, especially in this high-pace developing region?

Response 6: We understand the reviewer's concern about using year 2006 INTEX-B emission inventory for 2013. However, time dependent inventories are not available over this region. The comment has also been addressed in the response to the general comment of the reviewer,here extended:

As mentioned before, emission inventories over the South Asian region are not available for each year. We agree that by the year 2013 emissions might have changed but in the absence of such data, research studies focussing on the region resort to using various recent emission inventories representative of a different year (for e.g., Kumar et al., 2012b; Kumar et al., 2015; Ghude et al., 2016; Ojha et al., 2016). Our work aims to investigate the importance of emission inventories. In the present study, using INTEX-B inventory also serves to examine changes in the emissions in recent years by comparing the newest inventories (HTAP and SEAC4RS) to it. Similar comparison between emission inventories was also carried out in the study by Amnuaylojaroen et al. (2014) in which simulations were carried out over South east Asia for the year 2008 using emission inventories RETRO (year 2000), INTEX-B (year 2006), MACCity (year 2010) and SEAC4RS (year 2012).

To investigate the effects of different emission inventories on modeled ozone, other factors, such as simulation year, have to be kept the same. Also as mentioned in the response to comment 5, a simulation conducted with HTAP inventory for year 2010 showed small differences in estimated model biases. This clearly indicates that changing the model simulation year would not affect our conclusions.

Comment 7: Page 6, lines 198-200: But how the comparison would make sense given that the emissions are from different years and are also different between different inventories?

Response 7: We agree that such a comparison has limitations but we would again like to emphasize that regional bottom up inventories are not available over South Asia for every year and that studies have to rely on global inventories (such as HTAP) or regional inventories from specifc experiments in the region (SEAC4RS, INTEX-B) available for a recent year. Therefore it is important to know how different are the modeled ozone levels among different emissions and how do they compare with limited observational data before using model results for calculations of budget, and impacts on human health and crop yield. (Also see the responses to your general comment and comment 6).

Comment 8: Page 6, line 204: No, that too much difference, I do not think the authors can use (2004 or before) ozone measurements to validate model simulations for years 2013 using emissions from different temporal periods?? I think the authors need to reconsider all these comparisons.

Response 8: We agree that the observations at three stations are relatively old but excluding them doesn't change our conclusions (neither region wise nor for the domain).

We wish to keep these sites as this provides qualitative (if not quantitative) insight as to how model performs at these sites in terms of reproducing diurnal patterns. Datasets older than 10 years or more has been used in a previous study (Kumar et al., 2012b), however, we use more recent datasets, in general. We hope that reviewer would agree with our decision.

Comment 9: Page 6, lines 219-220: Could the authors provide quantitative numbers for this similarity between HTAP, INTEX and S4RS scenarios (e.g., r^2)? To me, they look quantitatively different.

Response 9: The quantitative assessment of similarity in simulated surface ozone among the three simulations is provided in the following table for both 24 h average and noontime (1130-

1630 IST) average at all grids in the domain. It's apparent from the variance of the residual that the scatter is relatively less for 24 h average indicating that the differences are smaller as compared to noontime averages. This information is now added in the revised manuscript (Page: 13 ; Lines: 481-483 and 484-486 ; supplementary material, Table. S5)

Table. Quantitative assessment of similarity between HTAP-RAMD2, INTEX-RADM2 and SEAC4RS-RADM2 scenarios for 24 h average and noontime (1130-1630 IST) average for simulated surface ozone mixing ratios

24 h average	**HTAP-RADM2 (a) vs INTEX-RADM2 (b)**	**HTAP-RADM2 (a) vs S4RS-RADM2 (b)**	**INTEX-RADM2 (a) vs S4RS-RADM2 (b)**
r^2	0.98	0.98	0.99
variance of the residual (b-a)	4.61	5.32	2.05
Noontime average			
r^2	0.96	0.96	0.98
variance of the residual (b-a)	18.26	21.24	11.70

Comment 10: Page 7, lines 241-250: Again, it is important to address here if the differences in the ozone production rates between different emission scenarios are related to using different temporal periods for the emission inventories or related to different emission inventories as it appears here?

Response 10: We are also trying to convey that in the absence of continuous bottom up regional emission inventory in this part of the world, studies analysing budget or impacts of ozone (typically using one of the inventories) should consider how results could have been different if another emission inventory (or model chemistry) would have been used. It is crucial to know the uncertainities associated with these results. While there have been numerous studies analysing processes, budgets and impacts, no comprehensive inter-comparison is available and we are here filling that gap. Nevertheless, we agree and now explicitly mention that more efforts are to be made to prepare high-resolution regional anthropogenic emissions over South Asia (Page:1 ; Lines: 34-36 ; Page: 14 ; Lines: 546-547).

Comment 11: Page 8, lines 304-318: So, are these differences related to chemical mechanism, or the constrained different overhead ozone column, or photolysis rates (Fast-J vs F-TUV) or different aerosol modules (static vs dynmic)?

Response 11: This comment has been responded previously (see response to comment 4) and mentioned again here. The major differences between two chemical mechanisms are due to differing inorganic reaction rates, while the effect of different photolysis schemes is small (Page: 11; Lines: 394-403 in the revised manuscript; also see Mar et al, 2016). Moreover, as the aerosol

radiation feedback is turned off, the observed differences are mainly result of differing gas phase chemistry. This is discussed and clarified in the revised version (Page: 11; Lines: 403-405).

Comment 12: Page 11, lines 403-406: The authors claim interesting similar results despite the use of different temporal emission, but I think that shows only possible compensating effects that lead to the claimed similar results despite different emissions… I think that the authors should seriously address this issue as it significantly affect the credibility of the results.

Response 12: We do not see the credibility of the results compromised, as we are trying to convey that the use of one of the available inventories arbitrarily would produce significantly different ozone fields and that the most recent inventory (SEAC4RS) coupled to RADM2 chemical mechanism is closer to the observational data from recent years.

According to the lines of the reviewer, we only said that it is interesting that model biases are similar between SEAC4RS and INTEX-B inventories, which were prepared for different time periods. The time periods as well as the input amount of emissions is explicitly given (see the referred statement and Table 2).

It is not possible to simply scale the emissions for difference in the time periods. For example, total NMVOC emissions were 26 million mol h^{-1} in the year 2006 (INTEX-B), 38.7 million mol h^{-1} in 2010 (HTAP) and 28.3 million mol h^{-1} in 2012 (SEAC4RS). Therefore one can not simply deduce a trend and scale the emissions, instead the emissions need to be prepared by taking an account of activity data on yearly basis in this region.

We have concluded that the most recent SECA4RS inventory coupled to RADM2 chemical mechanism is best suited inventory for simulating ozone fields over Indian region. The sentences referred to are suitably modified in the revised version (Page: 12; Lines: 462-463).

Comment 13: Page 11, 420: Again, I still not convinced by the "overall agreement", given that the model is constrained to emissions from different temporal periods than the measurements as well as the model simulations (using reanalysis products from year 2013).

Response 13: To summarise again, regional inventories are not available over the South Asian region for every year so air quality studies have to rely on emission inventories representative of a different year (for e.g., Kumar et al., 2012b; Kumar et al., 2015; Ghude et al., 2016; Ojha et al., 2016). One of our goals is to convey the uncertainties that can arise in ozone mixing ratio prediction due to choice of inventory (and also the employed chemical mechanism).

We agree with reviewer's opinion and are also trying to highlight through this work that the ozone observational network is to be further expanded and data to be archived, TOAR being one of such initiative (http://toar-data.fz-juelich.de/). While previous studies used much older observations, we incorporated new data especially over the rapidly changing Delhi region (and also Pune), having the same temporal period as the model run. Observations at Thumba and Jabalpur are also for the same year as the model. For other stations too we preferably used recent data (2009-2013). This information and limitations are discussed in detail in the paper (Page: 6, Lines: 226-234). Also as mentioned in a previous response, changing the model

reanalysis year doesn't impact the results, which we show in the paper by conducting dedicated numerical experiments (Fig. S4).

References

Amnuaylojaroen, T., Barth, M. C., Emmons, L. K., Carmichael, G. R., Kreasuwun, J., Prasitwattanaseree, S., and Chantara, S.: Effect of different emission inventories on modeled ozone and carbon monoxide in Southeast Asia, Atmos. Chem. Phys., 14, 12983-13012, doi:10.5194/acp-14-12983-2014, 2014.

Ansari, T. U., Ojha, N., Chandrasekar, R., Balaji, C., Singh, N. and Gunthe, S. S.: Competing impact of anthropogenic emissions and meteorology on the distribution of trace gases over Indian region, J. Atmos. Chem., doi:10.1007/s10874-016-9331-y, 2016.

Ghude, S. D., Jena, C., Chate, D. M., Beig, G., Pfister, G. G., Kumar, R., and Ramanathan, V.: Reduction in India's crop yield due to ozone, Geophys. Res. Lett., 41, 51971, doi:10.1002/2014GL060930, 2014.

Ghude, S. D., Chate, D. M., Jena, C., Beig, G., Kumar, R., Barth, M. C., PFister, G. G., Fadnavis, S. and Pithani, P.: Premature mortality in India due to PM2.5 and ozone exposure, Geophysical Research Letters, 4650 – 4658, 2016.

Girach, I. A., Ojha, N., Nair, P. R., Pozzer, A., Tiwari, Y. K., Kumar, K. R., and Lelieveld, J.: Variations in O3, CO, and CH4 over the Bay of Bengal during the summer monsoon season: shipborne measurements and model simulations, Atmos. Chem. Phys., 17, 257-275, doi:10.5194/acp-17-257-2017, 2017.

Hilboll, A., Richter, A., and Burrows, J. P.: NO2 pollution over India observed from space – the impact of rapid economic growth, and a recent decline, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2017-101, in review, 2017.

Jena, C., Ghude, S. D., Pfister, G. G., Chate, D. M., Kumar, R., Beig, G., Surendran, D., Fadnavis, S. and Lal, D. M.: Influence of springtime biomass burning emissions in South Asia on regional ozone: A model based case study, Atmos. Environ., 100, 37–47, doi:10.1016/j.atmosenv.2014.10.027, 2014.

Kumar, R., Naja, M., Pfister, G. G., Barth, M. C., Wiedinmyer, C., and Brasseur, G. P.: Simulations over South Asia using the Weather Research and Forecasting model with Chemistry (WRF-Chem): chemistry evaluation and initial results, Geoscientific Model Development 5, 619-648, 2012b.

Kumar, R., Barth, M. C., Pfister, G. G., Nair, V. S., Ghude, S. D., and Ojha, N.: What controls the seasonal cycle of black carbon aerosols in India?, J. Geophys. Res. Atmos., 120, 7788–7812, doi:10.1002/2015JD023298, 2015.

Mar, K. A., Ojha, N., Pozzer, A., and Butler, T. M.: Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: model evaluation and chemical mechanism comparison, Geosci. Model Dev., 9, 3699-3728, doi:10.5194/gmd-9-3699-2016, 2016.

Ojha, N., Pozzer, A., Rauthe-Schöch, A., Baker, A. K., Yoon, J., Brenninkmeijer, C. A. M., and Lelieveld, J.: Ozone and carbon monoxide over India during the summer monsoon: regional emissions and transport, Atmos. Chem. Phys., 16, 3013-3032, doi:10.5194/acp-16-3013-2016, 2016.