Response to Reviewer 2

Reviewer's comments are in italics, followed by our responses in non-italics.

This paper provides valuable insights for modeling NO_2 in south and east Asia. A comprehensive comparison between UKCA modeled NO_2 concentrations and OMI NO_2 observations and trend analyses using both datasets are performed. The results show that in south and east Asia, UKCA tends to overestimate NO_2 compared to OMI while exhibits similar trends over 2005-2015. However, more detailed and quantitative discussions need to be done to better illustrate the results.

Major comments:

(1) To me, the motivation of this study remains unclear. Does this paper aim at model evaluation? If so, I recommend a review of CCMs in the introduction, and more technical details of UKCA in section 2.1. Also, more background information is needed to explain why this paper focuses on south/east Asia besides the large population, e.g., lack of ground monitoring networks, and discrepancies between model and observations.

We agree we need to clarify the paper's motivation. Its motivations include: (i) to describe some of the difficulties in comparing column NO_2 in an atmospheric chemistry model with satellite measurements; (ii) to present such a comparison using the UKCA model; and (iii) to speculate on some of the atmospheric chemical processes that may contribute to model-measurement differences.

We take the point that we should clarify if what we are presenting is a model evaluation. We attempt to compare, as best we can, modelled NO_2 columns (and trends) with those measured by OMI, to try and draw conclusions about how well the model is performing. We feel that could be described as model evaluation. However, it is only a partial evaluation of how well the model is representing NO_2: e.g., we do not additionally compare with other available measurements of NO_2 (e.g., from surface sites and aircraft campaigns), nor associated relevant measurements (e.g., NO_2 deposition fluxes, or photolysis rates, etc.). A more comprehensive assessment of oxidised nitrogen would be very useful, but we feel is beyond the scope of the present study.

Previous studies have presented comparisons between modelled and satellite column NO_2 and generally carry out these comparisons using sensible methodologies (i.e. accounting for averaging kernels and time of satellite overpass), however we have not previously seen a simple step-by-step explanation of such a methodology and why it is important.

Understanding what a satellite measures is crucial for evaluating modelled NO_2 and seeing why that model evaluation carries greater uncertainties at certain locations and times of year (i.e. at higher latitudes in winter).

The paper focuses on S/E Asia as this is where satellites measure the largest NO_2 columns associated with anthropogenic emissions, and there have been large trends in these emissions over recent years.

We now include a brief description of the representation of NO_2 in UKCA in the introduction and the model description section. It is difficult to generalise this to all CCMs as they represent atmospheric chemistry to different degrees, but the NO_2 chemistry in UKCA is probably typical.

(2) Overall, there is a lack of comparisons with previous studies, e.g., is overestimating NOx a common problem for CCMs; does this study improve any existing problems? Also, the role and uncertainties of nitrogen deposition should be included. Moreover, the discussion and

conclusions need to be more quantitative. The author mentions uncertainties in NOx emission inventories and the representation of heterogeneous chemistry without quantifying the uncertainties, as a result, it is difficult to draw any strong conclusions.

Archibald et al. (2020) briefly evaluate NO_2 columns in UKCA by comparing with OMI data. These authors use a similar version of the model to that used in our study. E.g., Figure 18 of Archibald et al. (2020) shows global maps of simulated 2005-2014 DJF and JJA tropospheric NO_2 columns, together with UKCA-OMI differences. These results highlight similar issues to those found in our study: NO_2 is overestimated in UKCA over polluted regions of Asia, particularly during winter. Averaged over the whole latitude band 30-60N, the model overestimates DJF NO_2 and underestimates JJA.

We agree it is difficult to draw strong conclusions as we have not performed any sensitivity experiments (e.g., changing emissions or varying model chemistry). We feel this is for follow-up studies; our study only attempts to document current model performance and speculate on potential reasons for model discrepancies.

(3) Trend analysis:
What is the statistical method used for trend analysis? Are annual trends deseasonalized? Please add a paragraph of methodology description in Section 2.
What are the statistical significance values? I suspect that a lot of regions do not have statistically significant trends.

We applied linear regression to calculate trends in NO_2 concentrations. The annual means of NO_2 were used for the analysis, which inherently removes seasonal variability, ensuring that the trends reflect long-term changes rather than seasonal fluctuations.
Statistical significance was tested at a 95% confidence level (α=0.05). The t-statistic for each trend was calculated as t=trend/SE, where SE is the standard error of the trend estimate. The critical t-value ($t_{critical}$) was obtained from the t-distribution using degrees of freedom (df=n-2, where n is the number of years). Trends were considered significant if $|t|>t_{critical}$, indicating that the trend is unlikely to have occurred by chance with 95% confidence. Grid-boxes with significant trends are indicated by crosses on the revised Figures 9 and 10 (below).
Not all regions exhibit statistically significant trends. We have added a paragraph in Section 2 detailing the methodology for trend analysis, including the significance testing procedure to enhance clarity.

Is 2011 double-counted in both 2005-2011 and 2011-2015 periods?

Data from 2011 is included in both the time periods analysed. We wouldn't describe this as double-counting, as this makes it sound like a mistake. The year 2011 is deliberately included in both periods.

(4) Please highlight the novelty of this study.

Nitrogen dioxide columns in the UKCA model have only been briefly evaluated previously (Archibald et al., 2020). This study extends that analysis, looking in detail at diurnal and seasonal variations, and trends over 2005-2015. This study highlights significant discrepancies between UKCA and OMI measurements, especially in winter over China, and suggests some reasons why the model is under-performing. We hope this provides a useful guide for future model development and further evaluation of NO_x, both in UKCA and potentially in other models. These model discrepancies are important to address, since NO_x is important for a wide range of environmental problems that models are used to investigate.

Minor comments:

Line 52 and line 296: NOx: subscript
Modified

Line 82: BLH affects NO_2 column surface NO_2
Modified

Figure 5: it would be better to show the 2005 – 2015 mean instead of just using 1-year data. Also, please plot the number of OMI NO2 pixels for each month.

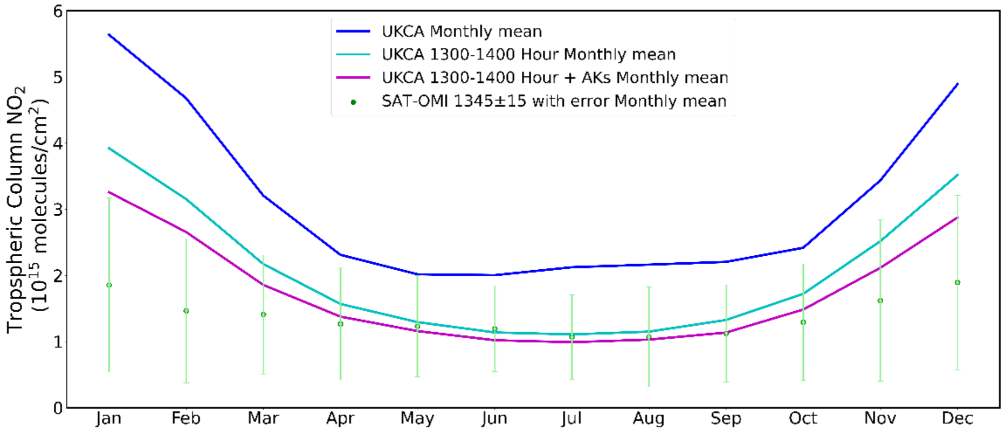
Thank you for your suggestion regarding Figure 5. We have revised the figure using the mean values for the years 2005–2015, as recommended (see below). However, regarding the number of OMI NO_2 pixels for each month, we encountered challenges in obtaining this data. The OMI pixels are irregular and follow a swath with varying sizes, making it difficult to compile consistent data over the ten years. The uncertainty bars partly reflect these variations in sampling, and we are confident that there are sufficient pixels for each month to allow us to perform a useful model-measurement comparison.

Figure 7 (a): please adjust the legend to avoid blocking the text

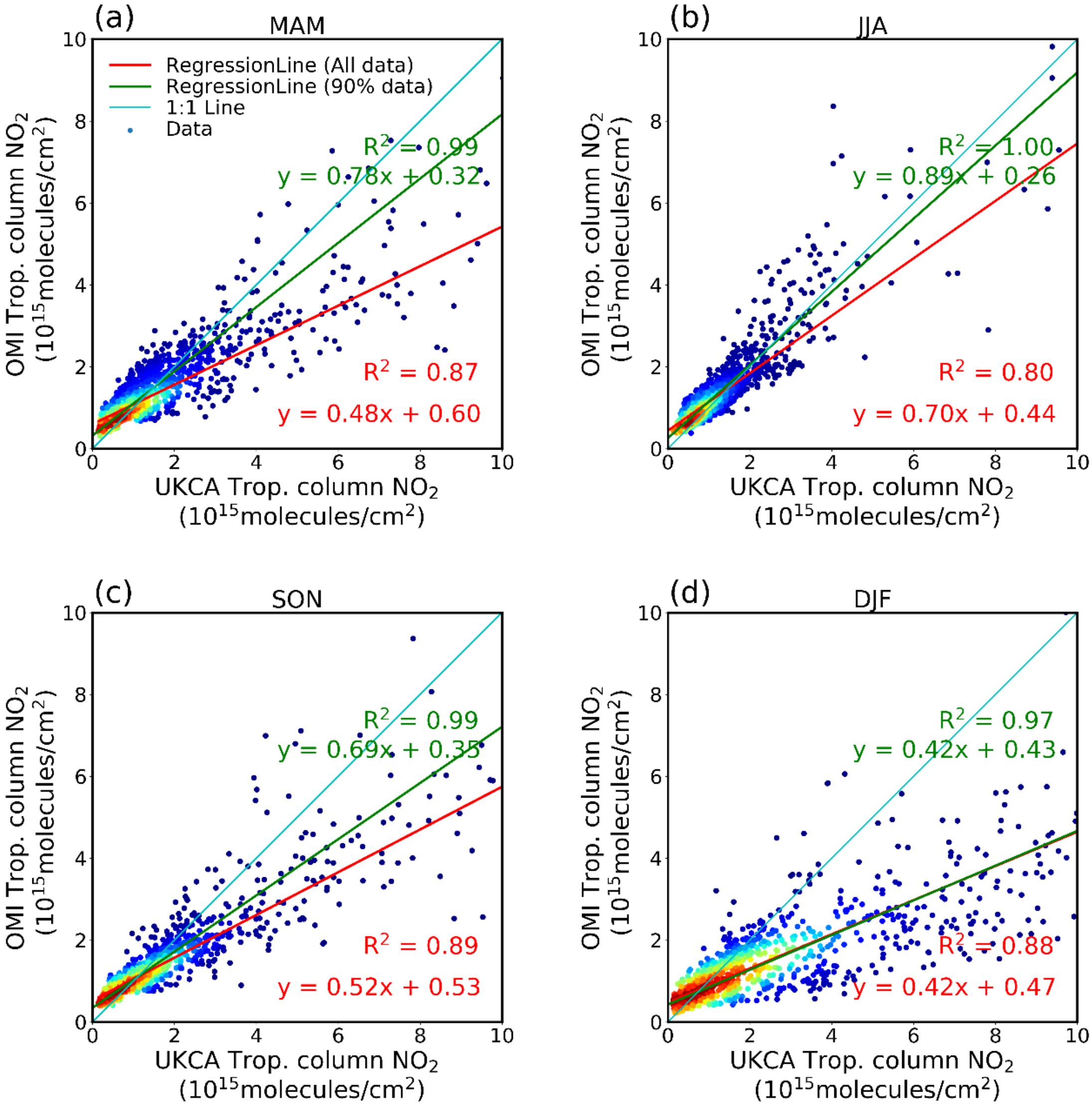
We have modified the legend in Figure 7(a) to prevent it from blocking the text. This adjustment enhances the overall readability of the figure.

Figure 11 and 12: add number of points to the scatter plot

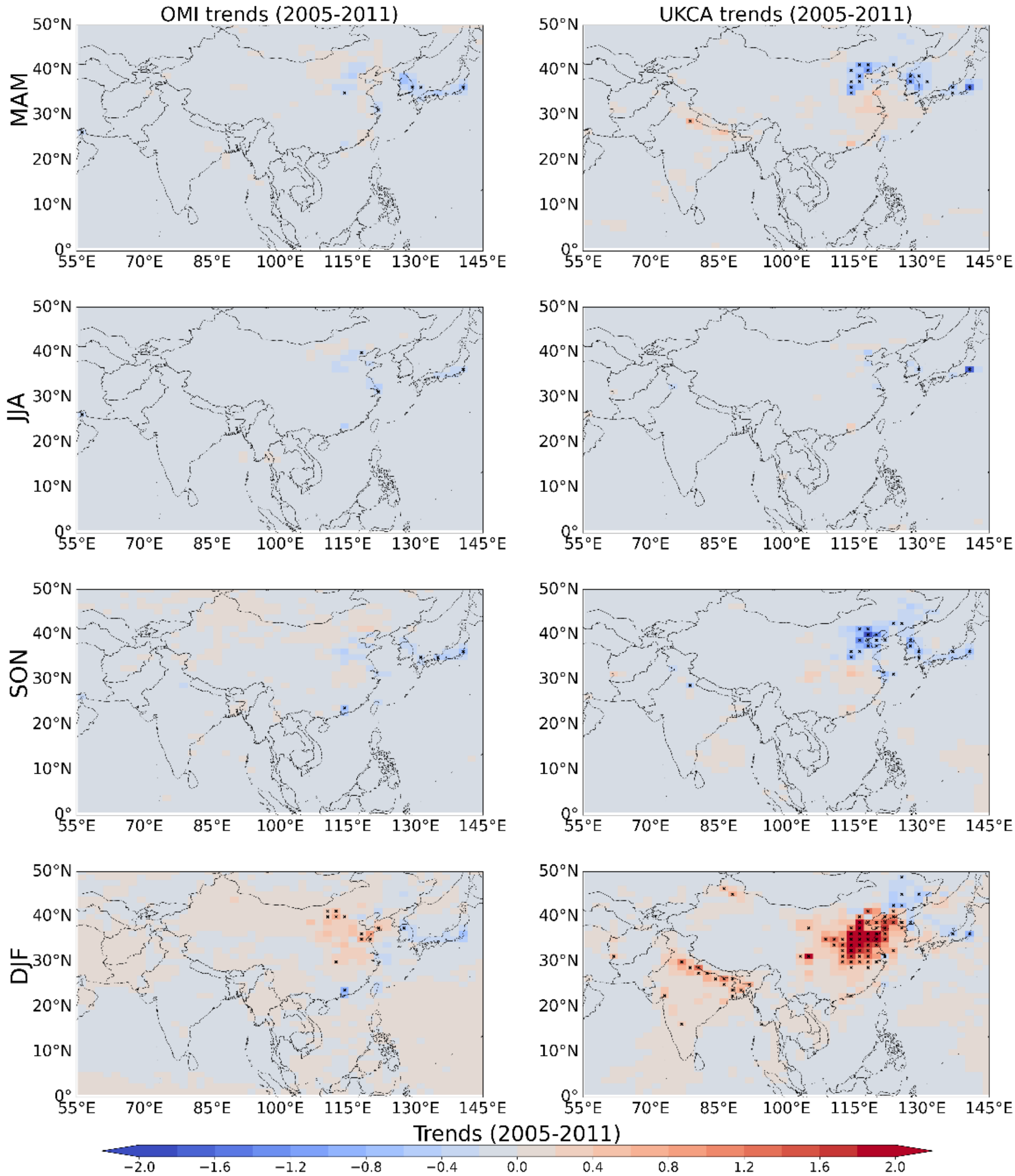
We have modified the figure captions to explicitly state that there are 2050 points used for each plot fit, while also acknowledging that some points are not on these plots as they fall outside the scales shown.



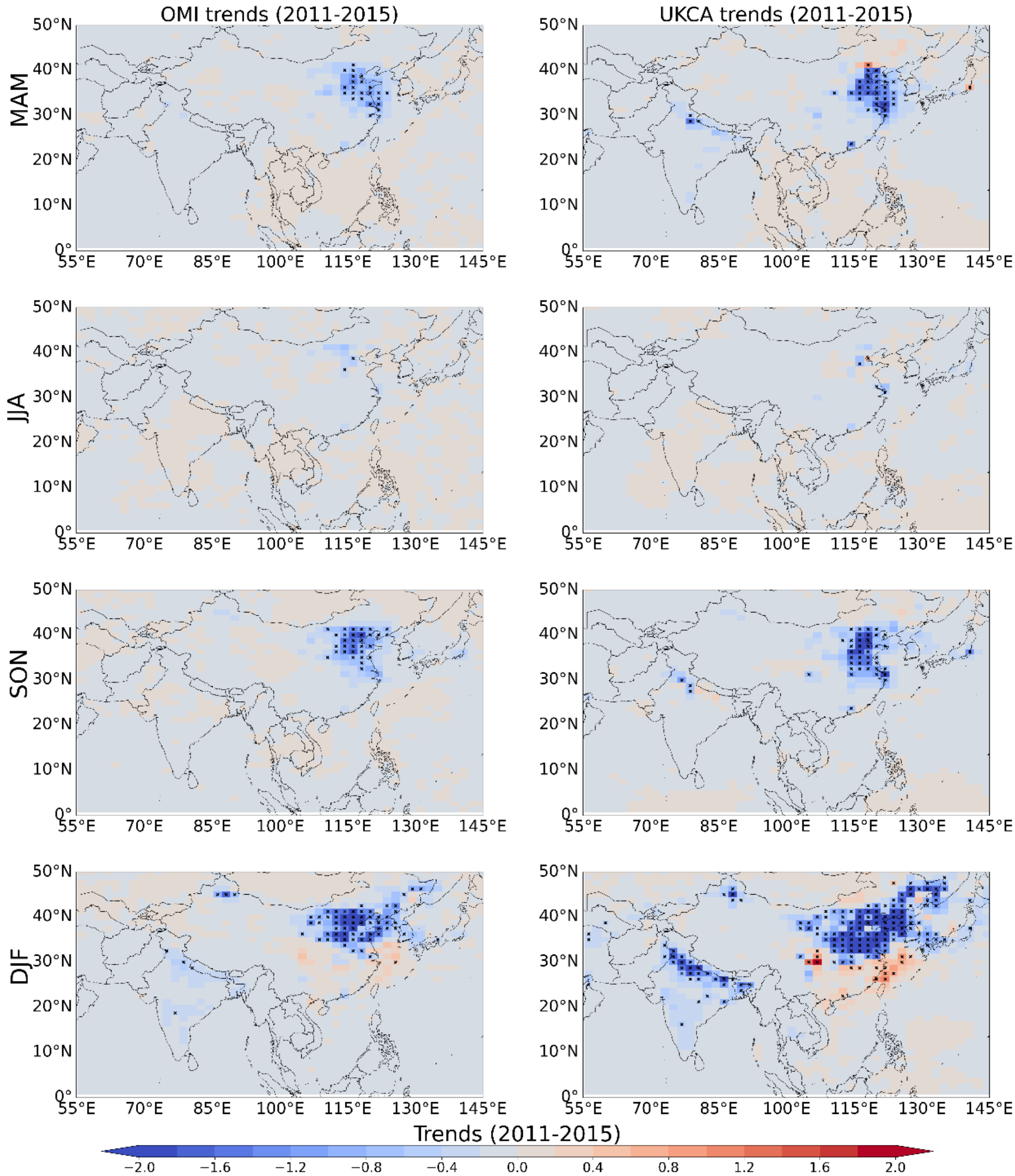
Revised Figure 5 Comparison of monthly mean tropospheric column NO_2, averaged over 2005-2015 for the whole S/E Asia region (Figure 1) from OMI (green, with uncertainty indicated by the bars), and from UKCA sampled in three different ways: (i) simple monthly mean (blue); (ii) sampled at the OMI overpass time (cyan); and (iii) sampled at the overpass time and with satellite averaging kernels applied (magenta).



Revised Figure 7 Scatter plots of OMI and UKCA tropospheric column NO_2 for the four seasons averaged over 2005-2015. Scatter data points are plotted as a heat map where red corresponds to more data. The 1:1 line is shown in cyan colour, best fit in red line (all data) and green line (lowest 90% of data). The equations of best fit and the coefficients of determination (R^2) are also shown in the respective colours.



Revised Figure 9 Trends of tropospheric column NO_2 (10^{15} molecules/cm^2/yr) from 2005 to 2011 from OMI (left) and UKCA (right) for the four seasons. Significant trends are indicated by crosses. Scatter plots of these data are shown in Figure 11. Equivalent data expressed as percentages are shown in Figure S8.



Revised Figure 10 Trends of tropospheric column NO_2 (10^{15} molecules/cm^2/yr) from 2011 to 2015 from OMI (left) and UKCA (right) for the four seasons. Significant trends are indicated by crosses. Scatter plots of these data are shown in Figure 12. Equivalent data expressed as percentages are shown in Figure S9.