RC2: 'Comment on egusphere-2024-570', Anonymous Referee #2, 07 Apr 2024

Edwards et al. present observations of diurnal variations in NO2 from GEMS, one of the geostationary satellites measuring the NO2 column at an hourly time scale. The authors quantify the diurnal variation using two metrics: the sum of absolute changes in the NO2 column over the day and the absolute deviation of the day's hourly observed NO2 relative to the observation at TROPOMI overpass time. The authors also utilize TROPOMI and Pandora measurements to help interpret the NO2 diurnal variation observed in the GEMS observations. The manuscript is well-written but requires some clarification on the context, such as the connections between satellite observations and modeling analysis. I suggest major revisions before publication in ACP.

We thank the Referee for their time in reviewing the paper. The comment above about the clarification on the context of the observations and modeling is addressed along with the other points in the answers below.

1. I suggest adding the equations to describe the matrices used to quantify NO2 diurnal variations. It is very hard to understand what exactly is defined as the "monthly average absolute change" in the figures.

Following the Referee's suggestion, the discussion of the diurnal change in NO2 in Section 3.2 has been significantly expanded and the equations used for calculation have been added to the text.

2. It is worth more discussion on the uncertainty in the GEMS NO2 column observations. What are the uncertainty levels of the hourly, daily, and monthly NO2 column measurements? Is the NO2 diurnal variation described in the paper sensitive to the measurement uncertainty?

The nominal GEMS pre-launch accuracy for the NO2 column is $1x10^{15}$ (molec.cm^{-2}) (Kim et al., BAMS 2020). The L2 product contains an error term for the NO2 spectral fit but this does not account for the AMF calculation error and the conversion to NO2 vertical column. Analysis for TROPOMI (van Geffen et al., 2022) identified the latter as the main source of error (around ±25 % over polluted regions). This comment has been added to the description of GEMS in Section 2.1 (#113). However, a positive bias in the GEMS V2 NO2 L2 data is often seen, especially over polluted areas, that has been characterized mainly through Pandora and other ground-based remote sensing comparisons of the type we report in Section 4. We also cite Kim et al. (2023) ACP SI who found a negative bias over clean areas attributable to representativeness error. Lange et al. (2024) (currently in review for the GEMS SI) compares GEMS to ground-based DOAS measurements and finds a median relative difference of +64 % and a correlation coefficient of 0.75. We have added this reference to Section 4 (#322). In short, there are known biases with this GEMS version that are being reported (as here) and that will hopefully be addressed in later data releases. Indeed, the biases were decreased between the Version 1 data and this Version 2

data as we discuss in Section 4. But the pattern of the diurnal variation is usually consistent between GEMS V2 and Pandora, between the different studies in this SI and can be reproduced in shape by the model. This is what we stress in this work - using these first GEO observations to assess the diurnal processes.

3. I can't follow the modeling method section and Section 5. How does the model compare against the GEMS observations? Is the model capturing the diurnal variations observed in the GEMS? If the model is biased, how do you utilize the model to gain a process-level understanding of the observed NO2 diurnal variation?

As discussed in answer to Referee's comment #2, there are known biases with this GEMS version that are being reported but the pattern of the diurnal variation that we investigate is usually consistent between GEMS, other measurements and models. We have added a new paragraph at the end of Section 2.4 (#171) and beginning of Section 5.1 to answer the comment about quantitative comparison between the GEMS data and model Section 5 (#357). This is also discussed in detail in answer to the last question of Referee#1.

4. The introduction section provides a detailed discussion of the satellite's capability of observing the NO2 column, but the result section delves directly into the diurnal variation in the NO2 column. I suggest restructuring the introduction section to highlight the significance of this study.

We have added more description at the beginning of the Introduction in Section 1 (#46) for the motivation for looking at the NO2 diurnal variation and GEMS being the first measurement from GEO to do this.

5. Figure 2: I don't think this figure is necessary for the main manuscript, you can move it to the supplement.

As the Referee questions below in comment #8, we needed to strengthen the connection between the GEMS data analysis and the Section 5 modeling work and discussion. We have now made it clearer in the modeling intro Section 2.4 (#167) that the diurnal emissions profile is used in Section 5.1 to investigate the sensitivity of the model TrC to the assumed emissions. These results are shown in Fig. 10 which are better explained by having the diurnal emissions profile in Fig. 2 as part of the main text. To make this clear, we also added a pointer to Fig. 2 in the presentation of the Base and Diurnal model simulations in Section 5.1 (#355).

6. Figure 9 and section 4: the comparison of the NO2 column between GEMS and other measurements at SEOUL-YN raises concerns about possible bias in GEMS measurement. Why there is a much steeper gradient in the observed NO2 column between 12 and 13 local times from GEMS? Why is it?

We agree that the Figures show a bias. As discussed in Referee's comment #2, there is a known positive bias in the GEMS V2 NO2 L2 data especially over polluted areas that has been characterized mainly through Pandora comparisons of the type we discuss in Section 4. The sharp gradient in GEMS data over Seoul-YN in June between 12:00 and 13:00 is due to the value at 13:00 being anomalously low. This occurs because of the limited number of measurements (2-3 per day) that meet the coincidence criteria with Pandora and then the low number of cloud-free days in the month. A note has been added to Section 4 (#339) to clarify this.

7. Following Figure 9, I wonder if the authors can point out and focus on regions where GEMS can provide more reliable NO2 column observations.

There are regions where the bias between the GEMS and Pandora is much smaller and these tend to be clean areas. Sometimes GEMS even shows a negative bias. But these are also regions where the diurnal cycle is much flatter and of less interest for understanding diurnal processes. We have seen this for our analysis of other Pandora sites, and it is also noted in Lange et al. (2024). We have added a note in Section 4 (#328).

8. Section 5: This part seems barely connected to other sections. More discussion is needed to strengthen the connections between sections.

Please see the answer to the last question of Referee#1 where this point is discussed in detail.