EGUSPHERE-2025-5285 – Author comments to RC4

https://doi.org/10.5194/egusphere-2025-5285

Legend: Referee comments in ***blue and italic***, author comments in **black**, changes in the manuscript text in **green**

Review #4

Bittner et al. present a valuable multi-year analysis of RGF using ground-based MAX-DOAS measurements at four sites: two biogenic and two anthropogenic. The author found that the patterns of the glyoxal-to-formaldehyde ratio (RGF) varied across the four study sites, including diurnal, weekly, and seasonal cycles. They also investigate how the temperature and NO2 drive the changes in RGF, and provide a valuable discussion of the four factors that reduce the comparability between different RGF values.

When the language is well used and data visualizations are very good, I feel the manuscript is very long, so I find it challenging to extract the take-home messages. Efforts should be made to improve the readability of the manuscript by combining certain sections and figures. More discussions should be included for other factors that can affect RGF, e.g., boundary layer height and air mass history. Machine learning could be a useful approach for this discussion. Overall, major revisions are required before the work can be published.

We would like to thank the Referee #4 for the valuable feedback that improved the manuscript significantly. We decided to adapt the proposed combined structure of the paper to make it easier to grasp. The results are now structured in (1) Temporal cycles (2) Link to meteorology (3) R_{GF}-NO_2 relationship (4) Comparability between R_{GF}. We further decided to remove the elevation section to make space for new content, as the elevation dependency was not easy to interpret (with and without O_4 correction) and R_{GF} did not show a clear signal. Furthermore, we computed the air mass history for each site (Fig. 4). Using backwards sensitivity runs with FLEXPART driven by ERA5 allows to visualize the footprints/sensitivities for all stations.

Detailed responses to the reviewer's comments can be found below. We hope that we have incorporated all suggestions and comments in a satisfactory way. At the end of this document, you will find a general section on additional changes we made during the review process

Major comments

1. *Section 3.2 and Section 3.5: The seasonal patterns are mainly attributed to temperature-driven HCHO changes (Figure 7), but other factors like boundary layer height, cloud coverage, and photolysis rates also need to be discussed.*

   See our response to reviewer 1 point 4.

2. *Section 3.6: The authors provide valuable empirical analysis of RGF-NO2 relationships across diverse environments. But there is a lack of mechanistic interpretation of how varying NOx regimes fundamentally alter CHOCHO and HCHO yields from different VOC precursors. In addition, it is unclear how the industrial processes in Incheon and Athens drive the observed difference in the CHOCHO-NO2*

   See our response to reviewer 1 point 1 for mechanistic interpretation and our response to reviewer 2 specific point 1 for an explanation regarding the emission sectors.

Minor comments

1. *Lines 58-86: The summary of the latest research on RGF is descriptive. It lacks clear motivation. I recommend that the authors make it more concise and compact for better readability.*

   We have rewritten the literature review to be more conceptual and concise, focusing on the scientific motivation rather than a descriptive list of studies. Lines 057-082: The ratio of glyoxal-to-formaldehyde (R_{GF}) was proposed by Wittrock et al. (2006) and Vrekoussis et al. (2010) as a potential proxy for differentiating VOC source types. Because CHOCHO and HCHO have similar sources and loss processes, subtle differences in VOC mixtures or source-specific yields are expected to be reflected in R_{GF}. The interpretation of R_{GF} as a diagnostic for VOC sources has remained inconsistent since its introduction. Vrekoussis et al. (2010) analysed two years of GOME-2 satellite data and found a strong spatial correlation between R_{GF} and VOC source categories, proposing a threshold of 4 % to distinguish anthropogenic sources (below) from biogenic or pyrogenic origins (above). They further observed decreasing R_{GF} with higher NO_2 levels and increasing values with greater vegetation density, quantified by the Enhanced Vegetation Index (EVI).

   Subsequent studies, however, produced mixed and sometimes contradictory results (Irie et al., 2011; DiGangi et al., 2012; MacDonald et al., 2012; Li et al., 2014; Chan Miller et al., 2014). Based on airborne in-situ data, Kaiser et al. (2015) shifted the focus toward VOC precursor speciation, finding that monoterpenes yield high R_{GF} values while isoprene yields low values. DiGangi et al. (2012) went further, proposing an interpretation opposite to that of Vrekoussis et al. (2010), with lower R_{GF} associated with biogenic sources and higher values with anthropogenic or pyrogenic origins. More recently, Chen et al. (2023) reported a positive correlation of R_{GF} with both EVI and NO_2 using TROPOMI data, and proposed that anthropogenic VOC emissions can be identified where R_{GF} > 4 % with additional constraints on EVI and HCHO columns. Hong et al. (2024) further argued that primary HCHO emissions bias $R_{GF,}$ and proposed the ratio of CHOCHO to secondary HCHO as a more reliable metric.

Further complexity was added by MAX-DOAS observations at rural and semi-urban sites in Southeast Asia. Hoque et al. (2018a, b) and Rawat et al. (2024) revealed pronounced seasonal and diurnal variability, while Xing et al. (2020) reported altitude-dependent changes in the diurnal cycle using vertical profile retrievals in China. Together, these studies found various influencing factors that contribute to the inconsistent results and highlight that the interpretation of R_{GF} remains challenging.

This study aims to systematically investigate the drivers and limitations of R_{GF} with the help of a multi-year, multi-site ground-based data set. MAX-DOAS observations from four sites in contrasting environments are analysed to investigate the overall magnitude of R_{GF}, temporal cycles (Sect. 3.1), link to meteorology (Sect. 3.2), and the R_{GF}--NO_2 relationship (Sect. 3.4). In addition, we identify and discuss four measurement-related effects in Sect. 3.4 that can hinder cross-study comparisons, with the aim of reassessing the suitability of R_{GF} as a proxy for VOC origin.

2. *Please include a table in Section 2.5 to summarize key characteristics: coordinates, type, instrument height, viewing direction, measurement period, and typical NO2 level for each site. This would help readers quickly grasp information about the four study sites.*

   In response to this excellent suggestion, we have added a new table (Table 1) that provides info on the four study sites.

3. *Figure 3: Please include the median and interquartile range for better statistics.*

   Figure 3 was reworked to better match the observations. (1) The median and IQR were added. (2) Values during the night were excluded to better resemble the conditions during the measurements. (3) Precipitation is now summed per month/hour before computing the median, correcting a previous pre-processing error.

4. *Figure 4: The difference in mark sizes for data availability is not clear. The same applies to Figures 5-7 and other similar figures.*

   The marker size shows the relative data availability in each panel. The bin with the maximum number of observations is defined as 100% per panel. If there are fewer data points in one bin (e.g. morning hours) the marker is smaller (line plot) or the alpha is lower (box plot).

   To improve interpretability, the previous linear mapping of marker size to observation count has been replaced with a saturating function: marker size decreases rapidly below 30 % of the maximum number of observations, and more slowly above this threshold. The sentences introducing this convention have been revised accordingly. The updated text now reads (Lines 343--344): Throughout this study, marker size is scaled to the number of observations per bin; smaller markers therefore indicate reduced bin size. The detailed mapping of bin sizes is provided in Fig. A8.

5. *Figure 4: What are the uses of overpass times? There is no discussion in section 3.2.*

   The overpass times are discussed in a later section (3.4.3). To avoid an extra figure, they are already included in Fig. 5 (previously Fig. 4).

6. *Lines 266-275: It is good that the authors mentioned findings in light of the literature. However, I feel that the comparison in RGF between this study and the literature is lacking. The same applies to the Lines 324 -329 in Section 3.2 Seasonal cycle.*

   We have reworked and expanded the literature comparisons in both sections.

   Lines 375–383: A direct quantitative comparison with previous studies is complicated by methodological differences: whereas DiGangi et al. (2012) report in-situ point measurements and Hoque et al. (2018a, b) and Rawat et al. (2024) derive R_{GF} from VCDs, our R^*_{GF} is based on corrected dSCDs, which integrate over a slant light path and are therefore sensitive to a different effective measurement volume (Sect. 3.4.1). Despite this, the qualitative diurnal patterns are broadly consistent. The midday peak observed at Incheon is also reported for rural and semi-urban sites in Southeast Asia (Hoque et al., 2018a, b; Rawat et al., 2024). However, the occurrence of similar patterns across differently classified sites highlights a broader challenge in the literature: the lack of a uniform site categorisation complicates cross-study comparisons of R_{GF}. At our predominantly biogenic sites ATTO and Orléans, the diurnal cycle is comparatively flat, which is consistent with the weak diurnal variability reported by DiGangi et al. (2012) for high-altitude biogenic sites.

   Lines 445–451: As with the diurnal cycle, a direct quantitative comparison is complicated by the fact that previous studies derive R_{GF} from VCDs, whereas our R^*_{GF} is based on corrected dSCDs at the lowest elevation angles, which correspond to a different effective measurement volume (Sect. 3.4.1). With this caveat in mind, the seasonal pattern at our anthropogenically influenced stations resembles most closely the winter enhancement reported by Xing et al. (2025) for Guangzhou. Our absolute R^*_{GF} values are lower than those reported by Xing et al. (2025), which may partly reflect the difference in measurement volume (dSCD vs VCD) rather than a true difference in R_{GF}. At our more remote stations, the magnitude of R^*_{GF} is comparable to that reported by Hoque et al. (2018b) for Pantnagar, even though no progressive annual increase is observed like at Phimai.

7. *Lines 280 - 282: The manuscript reports substantial differences in RGF values between biogenic (2.2-3.1%pt) and anthropogenic sites (3.5-4.2%pt) (Line 280-282). However, no statistical test is performed to validate these differences. Given the considerable overlap in standard deviations (e.g., Athens 3.5±0.4%pt and Incheon 3.7±0.7%pt), statistical tests (e.g., ANOVA with post-hoc testing or non-parametric equivalents) should be applied to confirm whether the observed differences between sites are statistically significant beyond random variations. This is particularly important for supporting the central conclusion that*

RGF values systematically differ between environment types (Line 630). The same concerns on annual mean values of RGF (Line 331).

We decided to focus our testing on the annual values of RGF and therefore removed the mentioning of the daily averages from the diurnal cycle chapter. The new section 2.4 describes the applied methodology. Lines 210-220: To assess whether observed differences in mean values are caused by random variability, we apply statistical tests in Sect. 3.1.2 and 3.1.3. Since measurements are available approximately every 30 minutes, consecutive data points may sample the same atmospheric event. To increase statistical independence, the data are temporally aggregated prior to testing. Where appropriate, a logarithmic transformation is applied to approximate normality.

To compare biogenic and anthropogenic environments, represented by ATTO + Orléans and Athens + Incheon, the data are aggregated to monthly means (e.g., two years of data yield 24 values). Differences between groups are tested using Welch's t-test (Welch, 1947; Delacre et al., 2017} applied to the log-transformed data, which accounts for unequal variances and sample sizes. The same aggregation strategy is used for station-to-station comparisons. In this case, a Welch analysis of variance (ANOVA) (Welch, 1951; Delacre et al., 2019) is applied first to the log-transformed data to assess overall differences among stations. It is followed by a Games—Howell post hoc test (Games and Howell, 1976), which evaluates pairwise differences while accounting for unequal variances and sample sizes.

When discussing the average values between environments, the text now reads (Lines 409-416): Aggregating all data points by month and grouping them by dominant environment, i.e. Orléans and ATTO as biogenic and Athens and Incheon as anthropogenic, yields mean R^*_{GF} values of 3.2+-1.1 % in the biogenic environment and 4.2+-0.8 % in the anthropogenic environment. Looking at mean R^*_{GF} per station leads to 3.4+-0.9 %, 2.7+-1.3 %, 3.9+-0.8 %, 4.6+-0.7 % for ATTO, Orléans, Athens, and Incheon respectively. Applying statistical tests, as described in Sect. 2.4, leads to significant differences (t=-5.8, p=8*10^{-8}) between the biogenic and anthropogenic group. A Welch-ANOVA (F = 19, p = 3*10^{-8}) combined with a Games-Howell post-hoc test resulted in significant differences for all station pairs except ATTO—Orléans and Athens—Incheon. More detailed results can be found in Table A6 + A7. It should be noted, that the aggregated data points maintain a significant autocorrelation due to the seasonal cycle.

8. *Figure 5 and Lines 296-297: It is not clear what information the authors want to convey with the O4*

We use the O_4 ratio to correct for physical differences in the retrieval. Since R^*_{GF} depends on the O_4 ratio (Eq. 3), we include it in the plots to show its contribution.

9. *Figures 4 and 5 can be potentially combined into one plot. The same applies to Figures 6 and 7 and also other figures.*

   We have combined the respective figures to one for every section.

10. *Lines 301-302: It is unclear how the dilution effects contribute to the observed shape.*

    The original idea was that high HCHO and CHOCHO levels in the morning/evening and low levels over noon might be partly explained by sampling the planetary boundary layer (PBL) differently due to PBL growth (scattering point inside/above PBL). But we are not able to conceptualize it without multiple assumptions and geometric simplifications. Therefore, we decided to remove this statement to avoid speculation.

11. *Section 3.7: The discussion on the factors affecting RGF values between different measurement techniques is highly appreciated. It would also be good to include suggestions to guide future work, e.g., "Future work should quantify these effects through radiative transfer simulations or direct instrument intercomparisons."*

    We have added this sentence to the conclusion paragraph discussing future work. Line 711: Future work should quantify these effects through radiative transfer simulations or direct instrument intercomparisons.

Technical comments

1. *Abstract: When mentioning HCHO and CHOCHO, it will be good to include their names.*

   Replaced their structure formulas with their full names. Lines 7,8,11

2. *The use of NMVOC is unnecessary, as it is used only four times. Instead, using non-methane VOCs is fine.*

   Replaced NMVOC with non-methane VOCs.

3. *Line 90: When "with meteorological variables" is mentioned, only temperature is analyzed throughout this study.*

   See our response to reviewer 1 point 4.

4. *Line 275: "to" is repeated.*

   Removed extra to.

5. *Do "%" and "%pt." mean the same thing?*

   No, as R_{GF} is given in % it complicates talking about absolute and relative differences of R_{GF}. Relative changes use % and absolute changes use %pt. For example (Line 172): For both stations, the mean weekday R_{GF} exceeds the mean weekend value by 0.5 %pt., corresponding to a reduction of approximately 10 % on weekends.

We have added an extra sentence to clarify this point (Line 146): Changes in R^*_{GF} are expressed in % for relative changes and in %pt. for absolute changes.

Other changes made to the manuscript during the review process:

- Every figure changed slightly except the map.
  - Removing the prebinning changed the numbers marginally (reviewer 3 specific point 2).
  - Changed to more outlier robust linear fit for Fig. 1, changed the O4 scaling factor for O4 vis dSCD in ATTO from 0.502 to 0.519.
- Added acknowledgements
  - Added missing persons and teams for supporting the stations
- Corrected wrong end year for Orleans: Figures already included two years of data (Jul 2023 until Jul 2025), whereas the text only stated until Jul 2024.
- Inclusion of backward sensitivity runs to get a better understanding of air mass history/station footprint with FLEXPART for reviewer comments (reviewer 4, reviewer 3 specific point 11, reviewer 2 general point 2, reviewer 1 point 12). Alexandros P. Poulidis who conducted the runs is now included in the co-authors.
- Restructured some sections for clarity. Removed introductory paragraphs and descriptive sentences to reduce bloat.
- Added figure showing diurnal cycles for different seasons (Fig A1).
- Removed ATTO O_4 scaling table and the other scatter plots used to determine ATTO O_4 vis scaling factor from the Appendix.
- Move most additional figures to the supplement material. Only two figures (Fig A1 & Fig A2) should stay in the appendix.