Reassessment of the glyoxal-to-formaldehyde ratio R_{GF} as a proxy for VOC source identification

Simon Bittner[1], Andreas Richter[1], Bianca Zilker[1], Sebastian Donner[2], Thomas Wagner[2], Alexandros P. Poulidis[1], Leonardo M. A. Alvarado[3], and Mihalis Vrekoussis[1,4,5]

[1]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany
[2]Max Planck Institute for Chemistry, Mainz, Germany
[3]German Aerospace Center (DLR), Earth Observation Center (EOC), Wessling, Germany
[4]Center of Marine Environmental Sciences (MARUM), University of Bremen, Bremen, Germany
[5]Climate and Atmosphere Research Center (CARE-C), The Cyprus Institute, Nicosia, Cyprus

Correspondence: Simon Bittner (simon.bittner@uni-bremen.de)

Abstract. The glyoxal-to-formaldehyde ratio (R_{GF}) has been proposed as a proxy to distinguish sources of volatile organic compounds (VOCs) in the atmosphere. However, the interpretation of its variability remains uncertain because of the diverse processes that affect VOC emissions and chemistry. In this study, we revisit the applicability and limitations of R_{GF} using multi-year ground-based MAX-DOAS measurements at four distinct sites: two biogenic (Orléans, France, and ATTO Tower, Brazil) and two anthropogenic (Athens, Greece, and Incheon, South Korea).

The results show higher R_{GF} in anthropogenic environments and lower at biogenic sites. Seasonal R_{GF} patterns are broadly ~~similar~~ consistent across sites, with ~~reduced~~ lower values in summer and ~~enhanced~~ higher values in winter, driven by ~~HCHO~~ formaldehyde variability. Diurnal cycles ~~, caused by CHOCHO variability ,~~ are primarily controlled by glyoxal variability and are more pronounced at urban sites, ~~with weekend effects~~ which also show a weekend reduction of 10 %. Correlations between R_{GF} and $\mathrm{NO_2}$ vary, even among anthropogenic stations, ~~indicating~~ highlighting the importance of local emission contributions.

~~Additionally, our analysis shows that increasing temperatures leads to a decrease of~~ Increasing temperatures from 15 to 35 °C decrease R_{GF} by up to 1.9 percentage points across all sites, ~~due to the more rapid increase of HCHO levels with temperature than CHOCHO.~~

~~Moreover, we~~ driven by the stronger temperature response of formaldehyde compared to glyoxal. We further discuss four effects that ~~reduce the comparability between different~~ complicate cross-study comparability of R_{GF} ~~values:~~ : differences in measurement volume, vertical sensitivity, ~~time dependence~~temporal sampling, and the impact of averaging-ratioing order.

Our findings suggest that ground-based remote sensing R_{GF} contains valuable diagnostic information about VOC source environments. However, its use as a universal proxy remains challenging, as our incomplete understanding of the various effects currently limits the reliable use of R_{GF} for VOC source attribution.

1 Introduction

The European Environmental Agency reported in 2024 that meeting World Health Organization (WHO) air quality standards across EU Member States could prevent 239000 annual deaths from fine particulate matter ($PM_{2.5}$), 70000 from tropospheric ozone (O_3), and 48000 from nitrogen dioxide (NO_2) exposure (European Environment Agency, 2024). Globally, the situation is comparable, with particularly high numbers of premature deaths occurring in Asia (Lelieveld et al., 2015).

~~Among the three pollutants, tropospheric~~ Tropospheric O_3~~—~~, which has strongly enhanced concentrations in summer smog~~—~~, is associated with increased cardiovascular and respiratory mortality (Bell et al., 2004; Turner et al., 2016). Its formation requires two precursors in the presence of sunlight: nitrogen oxides ($NO_x = NO_2 + NO$) and volatile organic compounds (VOCs) (Haagen-Smit, 1952). Understanding the role of these individual components is essential for effective ozone mitigation strategies. NO_x emissions originate primarily from fossil fuel combustion, followed by natural sources such as biomass burning, soil emissions, and lightning (Ehhalt et al., 2001; Seinfeld and Pandis, 2006).

~~In contrast, investigating~~ Investigating the origin of VOCs, focussing on non-methane VOCs~~(NMVOCs)~~, is more challenging, as they encompass a large and diverse group of compounds. In addition to their role in tropospheric ozone formation, they contribute to the formation of secondary organic aerosols (SOA) (Hallquist et al., 2009; Derwent et al., 2010) and cloud condensation nuclei (Zheng et al., 2020; Liu and Matsui, 2022). Their sources are generally categorised as biogenic, pyrogenic, or anthropogenic (Vrekoussis et al., 2010).

Among these categories, biogenic VOC emissions represent the largest share of total VOC emissions (Guenther et al., 1995; Stavrakou et al., 2009b). Vegetation emits up to 10000 different VOCs (Goldstein and Galbally, 2007), which are involved in a wide range of processes, including growth, development, communication, and defence against herbivores. Isoprene (C_5H_8) is the most commonly emitted VOC species, followed by monoterpenes ($C_{10}H_{16}$). Emission rates are influenced by many factors and vary across plant species, plant parts, and even leaf age (Laothawornkitkul et al., 2009; Zhang et al., 2023).

Another significant share of VOC emissions originates from pyrogenic sources, such as biomass burning. The combustion of biogenic material releases a complex mixture of species into the atmosphere, including a wide variety of VOCs. The composition of these emissions strongly depends on the material being burned (Gilman et al., 2015) and on moisture content (Paris et al., 2022).

Anthropogenic VOCs are emitted by a variety of sources. The Community Emissions Data System (CEDS) inventory indicates that energy production, road transportation, residential activities, and solvent usage are the dominant processes/sectors on a global scale (McDuffie et al., 2020).

Among VOC species, glyoxal ($CHOCHO$) and formaldehyde ($HCHO$) are key intermediate products of VOC oxidation in the atmosphere (Fu et al., 2008; Chan Miller et al., 2016). $HCHO$ is the most abundant atmospheric aldehyde, with primary emissions from vehicle exhausts (Nelson et al., 2008) and biomass burning (Lee et al., 1998; Andreae and Merlet, 2001). Its main source, however, is formation through secondary production from VOC oxidation (Fortems-Cheiney et al., 2012) and methane (CH_4) oxidation, which determines its background levels (Franco et al., 2016). $HCHO$ is removed from the atmo-

sphere by photolysis, reaction with hydroxyl radicals (OH), and deposition (Stavrakou et al., 2009b). Its typical tropospheric lifetime around midday is about 3 hours (Dienhart et al., 2021).

Glyoxal, the smallest dicarbonyl compound, shares similar sources with HCHO: primary emissions from biomass burning (Zarzana et al., 2017, 2018) and biofuel use (Fu et al., 2008), as well as secondary formation via VOC oxidation. Primary glyoxal emissions are generally small compared to its secondary production (Stavrakou et al., 2009a; Silva et al., 2018). Its tropospheric lifetime is short, on the order of a few hours (Volkamer et al., 2007; Myriokefalitakis et al., 2008; Fu et al., 2008). Glyoxal is removed through photolysis, reactions with OH, and both dry and wet deposition (Myriokefalitakis et al., 2008), with an additional important sink via ~~secondary organic aerosol (SOA)~~ SOA formation (Stavrakou et al., 2009a).

The ratio of ~~glyoxal to formaldehyde,~~ 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}.

~~In their 2010 study, Vrekoussis et al.~~ 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 ~~(ground pixel size: 80 ×40)~~ and found a strong spatial correlation between R_{GF} and VOC source categories~~(anthropogenic, biogenic, and pyrogenic). They concluded that R_{GF} values below~~ , proposing a threshold of 4 % ~~indicate predominantly anthropogenic sources , whereas values above 4suggest~~ to distinguish anthropogenic sources (below) from biogenic or pyrogenic origins ~~. Consistently, they observed a decrease of~~ (above). They further observed decreasing R_{GF} with ~~increasing~~ higher NO_2 levels and ~~an increase with higher~~ increasing values with greater vegetation density, quantified by the Enhanced Vegetation Index (EVI).

~~Since the initial study, further investigations led to~~ Subsequent studies, however, produced mixed and sometimes contradictory ~~findings, making it difficult to form a consistent interpretation of R_{GF}. Several studies supported or challenged the earlier conclusions~~ results (Irie et al., 2011; DiGangi et al., 2012; MacDonald et al., 2012; Li et al., 2014; Chan Miller et al., 2014). ~~For example, DiGangi et al. (2012), based on~~ Based on airborne in-situ data~~from two campaigns, proposed an interpretation opposite to that of Vrekoussis et al. (2010): lower R_{GF} for biogenic sources and higher values for anthropogenic or pyrogenic origins.~~

~~In 2015, Kaiser et al. reassessed R_{GF} and~~ , Kaiser et al. (2015) shifted the focus toward VOC ~~precursors. Using airborne in-situ data, they found that R_{GF} primarily depends on the speciationof hydrocarbon precursors, with monoterpenes yielding~~ precursor speciation, finding that monoterpenes yield high R_{GF} values ~~and isoprene~~ while isoprene yields low values.

~~At one rural (Thailand) and one semi-urban site (India), Hoque et al. (2018a, b) and Rawat et al. (2024) used MAX-DOAS observations to investigate~~ DiGangi et al. (2012) went further, proposing an interpretation opposite to that of Vrekoussis et al. (2010) , with lower R_{GF} ~~. They reported higher values during the wet season, a diurnal cycle peaking at noon, and seasonal variability between sites, periods, and different platforms. Similarly, Xing et al. (2020) used two weeks of MAX-DOAS vertical profile data from Chongqing, China, and found altitude-dependent changes in the diurnal cycle of R_{GF}.~~

~~A recent satellite-based study by Chen et al. (2023) using TROPOMI data~~ 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 ~~. Their results support the latitude dependence noted earlier by Myriokefalitakis et al. (2008) and they~~

reported reduced R_{GF} values during the COVID-19 lockdown. Chen et al. (2023) suggested that regional identification of using TROPOMI data, and proposed that anthropogenic VOC emissions is feasible under the following conditions: can be identified where $R_{GF} > 4\%$ and EVI < 0.2, with additional constraints on EVI and HCHO columns $> 0.5 \times 10^{16}$.

Another study, by Hong et al. (2024) , using MAX-DOAS data from four megacities in China, proposed excluding primary . Hong et al. (2024) further argued that primary HCHO emissions from the analysis. They suggested that the different primary emissions between and reduce emissions bias R_{GF} accuracy to diagnose VOC sources. By excluding the primary with a multiple linear regression model, they propose , and proposed the ratio of CHOCHO to secondary HCHO as a more reliable metric at surface level.

In summary, although considerable progress has been made over the past 15 years, the 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 and its behaviour differs between platforms and datasets.

This study systematically investigates aims to systematically investigate the drivers and limitations of R_{GF} using MAX-DOAS data from four stations located in different environments . We analyse its magnitude with the help of a multi-year, temporal patterns, and correlations with meteorological variables.Moreover, we discuss four effects, that potentially hinder comparisons of different 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} to, in the end, enhance our understanding and reassess –NO_2 relationship (Sect. 3.3). 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 Methods and Datasetsdatasets

2.1 MAX-DOAS

Multi-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) is a remote sensing technique that uses scattered sunlight in the ultraviolet (UV) and visible (VISvis) spectral ranges to determine trace gas concentrations, integrated along the average atmospheric light path. By computing optical depth from the measured spectrum and a reference spectrum, and comparing it to the known absorption cross-sections of specific trace gases, their atmospheric abundance can be quantified. The spectral fitting process focuses on the differential absorption structures within absorber-specific wavelength intervals, known as fit windows (Hönninger et al., 2004; Platt and Stutz, 2008).

The term Multi-Axis refers to the instrument's ability to scan in multiple viewing directions. By measuring at various elevations (vertical) and azimuths (horizontal), different atmospheric layers can be probed. Observations at high elevation angles (around 90 °, known as zenith-sky direction) are used for stratospheric absorbers, while low-elevation, off-axis measurements

in various azimuth directions are more sensitive to boundary layer trace gas concentrations (Hönninger et al., 2004; Wittrock et al., 2004; Platt and Stutz, 2008).

The DOAS retrieval yields the measured slant column density ($\mathrm{SCD_{meas}}$), relative to a reference spectrum with its own SCD ($\mathrm{SCD_{ref}}$). ~~Because DOAS captures only the differential absorption between the measured and reference spectra, it provides the differential slant column density (dSCD), see Eq. (2).~~ Mathematically, the SCD is defined as the integral of the absorber number density (~~n_{meas} or n_{ref}~~n) along the effective light path (~~ds_{meas} or ds_{ref}~~ds) from the top of the atmosphere (TOA) to the ground, see Eq. (1). Because DOAS captures only the differential absorption between the measured and reference spectra, it provides the differential slant column density (dSCD), see Eq. (2).

$$\mathrm{dSCD} = \mathrm{SCD_{meas}} - \mathrm{SCD_{ref}} = \int_0^{TOA} n_{\mathrm{meas}}(s)\, ds \tag{1}$$

$$\mathrm{dSCD} = \mathrm{SCD_{meas}} - \int_0^{TOA} n \mathrm{SCD_{ref}}\, ds_{\mathrm{ref}}. \tag{2}$$

In this study, we ~~primarily~~ use off-axis measurements at low elevation angles ~~(~~from 1 °–3 °~~)~~. The atmospheric abundances are retrieved with sequential fits, where the reference spectrum is the corresponding zenith-sky measurement closest in time or interpolated to the measurement time. This setup has the advantage that most stratospheric influences and diurnal changes in viewing geometry cancel out, so that changes in the dSCD reflect enhancements of the trace gas in the boundary layer near the ground. Measurements at 30 ° viewing elevation, representing a geometric approximation of the vertical column density (VCD), are shown in the supplement (Figs. A13 and A14). However, the limited number of data points remaining after filtering, together with the reduced variability of CHOCHO, renders these data unsuitable for the present analysis.

The multi-year dataset used here ~~, comprises~~ stems from four stations in different environments: ATTO Tower (Brazil), Orléans (France), Athens (Greece), and Incheon (South Korea). ~~The stations are introduced again in Sect. 2.5; measurement periods and instrument alignments are listed in table A10.~~ Three of the four instruments (Athens, Orléans, and Incheon) were developed and deployed by the University of Bremen and therefore use identical fit settings for NO_2, CHOCHO, and HCHO. Measurements from ATTO were obtained using a different instrument developed and evaluated by the Max Planck Institute for Chemistry (MPIC) (Donner, 2024), and thus, different fit settings were applied. All fit settings of the Bremen instruments are listed in the supplementary material (Tables A1, A2, A3, and A4). The fit settings for the instrument evaluated by MPIC are given in Donner (2024) in Tables 9.1-9.5.

We apply several quality filters based on the root mean square (RMS < 0.001) of the fit residual, intensity (with separate thresholds for UV and vis per station), solar zenith angle (SZA ~~< 80°~~ < 80 °), and the relative slant column density error (~~< 50%~~ < 50 %). The relative error filter for the dSCDs constrains the propagated uncertainty of R_{GF} to below 71 % ~~. All applied~~but indirectly also filters out situations with low atmospheric concentrations. No clear-sky filtering is applied. All thresholds are summarized in the ~~appendix~~ supplement (Table A5). ~~To facilitate data merging with the meteorological data, the measurements are binned to 30-minute resolution.~~

2.2 Computation of R_{GF}

We calculate R_{GF} for each ~~30-minute bin using the~~ pair of corresponding quality-filtered CHOCHO and HCHO dSCDs. Since CHOCHO and HCHO are retrieved in different spectral ranges, atmospheric scattering processes, such as Rayleigh scattering, vary, resulting in different effective light path lengths (Seyler et al., 2017). This discrepancy can introduce systematic differences between the two dSCDs, as each trace gas effectively samples a slightly different part of the boundary layer.

To estimate and correct for differences in light path lengths, we use the collision-induced absorption of O_2–O_2, typically approximated as O_4, which must be included as ~~a pseudo-absorber~~ an absorber in DOAS retrievals (Finkenzeller and Volkamer, 2022). The vertical profile of O_4 is well characterised; its ~~vertical column density (VCD)~~ VCD can be accurately calculated because O_2 concentration decreases approximately exponentially with altitude, producing a known vertical distribution of O_4. ~~To reduce systematic differences between fits in the UV and the visible part of the spectrum, we~~ We apply a first-order correction by multiplying R_{GF} with the inverse of the O_4 dSCDs from the corresponding wavelength regions, see Eq. (3). This correction approach is effective because the O_4 VCD cancels out in the process, leaving only the ratio of the respective air mass factors (~~AMFs~~$\mathrm{AMF}^{\mathrm{O_4}}$, $\mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{ref}}$), which accounts for differences in physical processes, see Eq. (4). ~~In the equations, the AMF of the measured spectrum is shown as is ($\mathrm{AMF}^{\mathrm{O_4}}$), while the AMF from the reference spectrum is denoted as $\mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{ref}}$.~~

The O_4 correction ~~is valid under the assumption~~ assumes that the vertical profiles of CHOCHO and HCHO closely resemble ~~the profile~~ that of O_4, since the O_4 ~~air mass factor (AMF)~~ AMF is used to correct for differences in effective light path length (Sinreich et al., 2013). This assumption is reasonable for our dataset, ~~since~~ as we focus on the lowest elevation angles, where ~~the~~ slant columns are dominated by ~~absorption in the layer close to the ground. Enhancements~~ near-surface absorption. However, when the profiles of CHOCHO and HCHO deviate from the exponential O_4 profile, the accuracy of the correction decreases. Such deviations may arise, for example, from enhancements at elevated layers ~~, e.g. from~~ due to fire plumes, ~~would weaken this approach~~or from a box-shaped profile under conditions of strong atmospheric stratification.

$$R^{*}_{\mathrm{GF}} = \frac{\mathrm{dSCD}^{\mathrm{CHOCHO}}_{\mathrm{vis}}}{\mathrm{dSCD}^{\mathrm{HCHO}}_{\mathrm{UV}}} \cdot \frac{\mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{UV}}}{\mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{vis}}} \tag{3}$$

$$= R_{\mathrm{GF}} \cdot \frac{\cancel{\mathrm{VCD}^{\mathrm{O_4}}}(\mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{UV}} - \mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{UV,ref}})}{\cancel{\mathrm{VCD}^{\mathrm{O_4}}}(\mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{vis}} - \mathrm{AMF}^{\mathrm{O_4}}_{\mathrm{vis,ref}})} \tag{4}$$

Overall, the O_4 correction of R_{GF} is important for interpreting the results because the physical processes, influencing the effective light path, are systematically different in the spectral ranges of CHOCHO and HCHO. Since dSCDs are used to compute R_{GF}, if left uncorrected, these light path effects alter the values of R_{GF}, hiding the influence of the actual drivers. Figure 5 (bottom row) illustrates the impact of the correction. The O_4 ratio is relatively constant over the day and primarily reduces the overall magnitude of the corrected R_{GF}. Throughout this study, we denote the uncorrected glyoxal-to-formaldehyde ratio as R_{GF} and the O_4-corrected ratio as R^{*}_{GF}. Changes in R^{*}_{GF} are expressed in % for relative changes and in %pt. for absolute changes.

The O_4 dSCDs used for correction are obtained from the respective NO_2 fits in the visible and UV ranges for the Bremen and MPIC instruments. As the MPIC instrument ~~covers a shorter wavelength range, the vis fit window extends only until 460 and~~

therefore does not cover the O_4 absorption band at 477 nm, which reduces the quality of O_4 dSCDs in the visible . The ratio of is reduced. Comparing the O_4 dSCDs from UV and vis is shown dSCDs for all sites in Fig. 1for all sites. The , shows that the original data from ATTO (Fig. 1a) deviates from the other stations, which we primarily attribute . We primarily attribute the higher slope to the lower quality of the O_4 retrieval in the visible.

To remove this bias systematic bias for ATTO in our study, we scale the O_4 vis dSCDs at of ATTO by using the data from all other sites (Orl+Ath) as the reference. A slope is computed for the original dataset (m_{ATTO}), see , Fig. 1a), and the reference dataset (m_{ref}), see , Fig. 1c). Both slopes are then used to scale the O_4 vis dSCDs, see Eq. (5), which results in yields a more consistent behaviour for ATTO (Fig. 1b). This scaling is calculated and applied for all elevation angles individually, see Table A11. The effect of the scaling is, that the affects the magnitude of the O_4 ratio is and makes it comparable with the other stations, see sites (Fig. ??+??5m). A comparison of the original and scaled O_4 ratio for the other figures is found in the appendix supplement in Fig. A17.

$$\mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{vis,scaled}} = \mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{vis}} \cdot \frac{m_{\mathrm{ref}}}{m_{\mathrm{ATTO}}} \tag{5}$$

Overall, the correction of R_{GF} is important for interpreting the results because the physical processes, influencing the effective light path, are systematically different in the spectral ranges of and . Since dSCDs are used to compute R_{GF}, if left uncorrected, these light path effects alter the values of R_{GF}, hiding the influence of the actual drivers. Figure ?? (bottom row) illustrates the impact of the correction. The ratio is relatively constant over the day and primarily reduces the overall magnitude of the corrected R_{GF}. In the following sections, we denote the uncorrected glyoxal-to-formaldehyde ratio as R_{GF} and the ratio with the correction as R^*_{GF}.

2.3 Meteorological data

2.2.1 Uncertainties of R_{GF}

Uncertainties from MAX-DOAS data can be grouped in (1) random effects and (2) systematic effects. Following the error budget discussion from Pinardi et al. (2013) for the HCHO retrieval from MAX-DOAS data, random uncertainties are connected to photon shot noise for silicon based detectors and are generally well captured, if not even overestimated, in the dSCD uncertainties from the DOAS fit. For scientific grade instruments, the systematic uncertainties outweigh the random uncertainties. Pointing misalignments, uncertainties of the wavelength calibration, and the uncertainties in the retrieval are common sources for systematic uncertainties (Roscoe et al., 2010; Pinardi et al., 2013). These amount to around 20 % for HCHO (Pinardi et al., 2013) and are typically not included in dSCD uncertainties.

Systematic differences in data collection and processing between sites are unavoidable. The instruments do not share identical hardware, and ATTO, the only instrument operated by the Max Planck Institute for Chemistry, uses slightly different fit settings compared to the other instruments. In addition, the O_4 vis dSCDs at ATTO are scaled as described above. At Incheon, viewing elevations below 3 ° are blocked, thereby reducing sensitivity close to the surface compared to measurements at 1 ° or 2 ° elevation. At Athens, the instrument is located at 500 m above sea level, while the city centre lies near sea level. Under

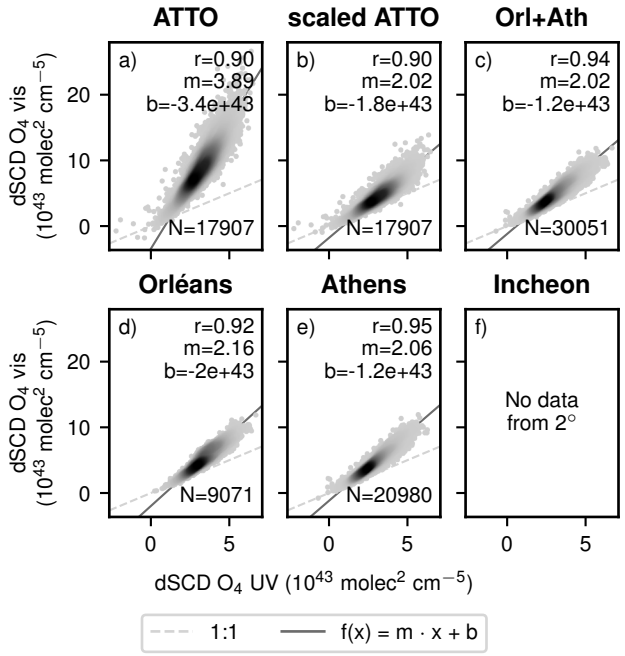

Figure 1. ~~Scatter plots showing the correlation of dSCDs from the UV wavelength range with the vis wavelength range~~O_4 vis dSCDs as a function of O_4 UV dSCDs for different datasets at 2 ° viewing elevation. The light grey dashed line indicates the 1-to-1 line and the gray solid line indicates a orthogonal linear fit with the specified parameters. The density of the data points is indicated by the hue, denser regions are shown in darker~~dark~~ grey.

shallow boundary layer conditions, such as in winter, the effective light path may therefore only partially sample the polluted boundary layer, resulting in lower measured columns. However, since CHOCHO and HCHO concentrations peak in summer, when the boundary layer is typically well developed, this effect is expected to be small.

To quantify the uncertainty of R_{GF}, we modify the uncertainty propagation from Vrekoussis et al. (2010) to include the O_4 ratio:

$$R^*_{\mathrm{GF}} = \frac{a}{b} \cdot \frac{c}{d} \tag{6}$$

$$s_{R^*_{\mathrm{GF}}} = \left[\left(\frac{c}{bd} \cdot s_a \right)^2 + \left(-\frac{ac}{b^2 d} \cdot s_b \right)^2 + \left(\frac{a}{bd} \cdot s_c \right)^2 + \left(-\frac{ac}{bd^2} \cdot s_d \right)^2 \right]^{\frac{1}{2}} \tag{7}$$

with a, b, c, d, s_x representing $\mathrm{dSCD}^{\mathrm{CHOCHO}}_{\mathrm{vis}}$, $\mathrm{dSCD}^{\mathrm{HCHO}}_{\mathrm{UV}}$, $\mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{UV}}$, $\mathrm{dSCD}^{\mathrm{O_4}}_{\mathrm{vis}}$, and the respective standard error. We use the uncertainties obtained from the DOAS fit to account for random uncertainties. The annual and seasonal median uncertainties of R^*_{GF} per station are listed in Table A9 in the supplement. The random uncertainties are higher during winter (Orléans,

Athens, Incheon) and wet season (ATTO). The relative uncertainties of R^*_{GF} on an annual scale range from 10 % to 20 % for all stations.

2.3 Auxiliary datasets

We use meteorological data to associate changes in R^*_{GF} with specific meteorological conditions, thereby extending our understanding of its driving factors. To ensure consistency across all stations throughout the measurement periods, we selected data from the ECMWF Reanalysis v5 (ERA5) dataset. ERA5 provides hourly gridded data (0.25° × 0.25° grid spacing). The meteorological variables included in this analysis are temperature at 2 m, dew point temperature at 2 m, boundary layer height, short-wave radiation, total precipitation, and wind speed and direction at 100 m (Hersbach et al., 2023). To merge the ERA5 data with the MAX-DOAS datasets, the ERA5 datasets are interpolated in time ~~from hourly values to every 30 minutes.~~ to the timestamp of each measurement. Relative humidity is computed via the Magnus approximation from the temperature and dew point temperature.

2.4 ~~VOC emission data~~

To investigate differences in ~~anthropogenic VOC~~ emission sources between the sites~~Athens and Incheon~~, we use the CAMS-GLOB-ANT version 6.2 ~~emission dataset created by~~ (Soulie et al., 2024) and the CAMS-GLOB-BIO version 3.1 (Sindelarova et al., 2022) emission datasets created by the Copernicus Atmosphere Monitoring Service (CAMS) ~~(Soulie et al., 2024)~~ and provided by ECCAD (Granier et al., 2019). The ~~annually gridded product~~ data is used in Table 1 characterising the stations by their NOx to VOCs ratio and anthropogenic VOCs to biogenic VOCs ratio. In addition, the anthropogenic contributions of non-methane VOCs for both urban sites during the observations are used to aid interpretation of the R^*_{GF}–NO_2 relationship in Fig. 12. For the respective contributions, the annually gridded CAMS-GLOB-ANT (0.1° × 0.1° grid spacing) for ~~NMVOCs~~ non-methane VOCs and NOx (in Tg) and CAMS-GLOB-BIO (0.25° × 0.25° grid spacing) for all VOC species (in Tg) are summed up over a region enclosing the ~~two cities~~sites, see Fig. A3. ~~Afterwards the annual emissions are averaged over the years of our ground-based coverage for both cities: Athens (2021-2023), Incheon (2021, 2022).~~ Carbon monoxide, methane, methyl chloride, methyl iodide, methyl bromide, and hydrogen cyanide are excluded for the contributions of biogenic VOCs from CAMS-GLOB-BIO.

To quantify the sensitivity of our measurements to nearby source regions, we performed backward simulations with FLEXPART version 10.4 (Pisso et al., 2019), driven by ERA5 meteorological fields. Hourly footprints were generated by initiating one simulation per hour, each with an one-hour emission pulse and a three-day backward integration period. Residence times (i.e. sensitivity, in s) were integrated over the full atmospheric column and accumulated over the entire simulation period. The released tracer was configured as a proxy for CHOCHO with a lifetime of 3 h. For each station, we selected one representative year in which the annual wind rose closely matched the corresponding multi-year wind rose (see Fig. A18). As the aim of the simulations is to study the spatial distribution of the footprint, only normalised sensitivity with respect to the maximum value is studied here.

2.4 Statistical tests

To assess whether observed differences in mean values are caused by random variability, we apply statistical tests in Sect. 3.1.1 and 3.1.1. 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. To investigate a potential weekend effect, the data are aggregated to weekly means separated into workdays and weekends for each station. Welch's t-test is then applied to the corresponding subsets.

2.5 Measurement sites

Four measurement sites were selected according to their predominant environmental characteristics. Each site was classified based on its surroundings and the chosen viewing direction (Fig. 2 and Table 1). Athens and Incheon represent anthropogenic environments due to enhanced NO_2 levels (Mavroidis and Ilia, 2012; Nguyen et al., 2015; Gratsea et al., 2016; Lange et al., 2024) and high population density within their metropolitan areas (Kim et al., 2021; Hellenic Statistical Authority, 2024).

The third station, Orléans, is classified primarily as a biogenic environment. This classification is supported by relatively low observed median NO_2 levels, and a viewing direction aimed directly over forest canopies. The fourth station, ATTO, is similarly considered biogenic, given its remote location within the Amazon rainforest.

Potential pyrogenic influences at ATTO (wildfires during the dry season) and Athens (occasional wildfires) are neglected, as we expect such events to ~~occur only infrequently within~~ influence our measurements only infrequently during our measurement periods.

2.5.1 Athens

The instrument in Athens is located at the National Observatory of Athens in Penteli, Greece. The Athens metropolitan area, with approximately 3 million inhabitants in the Attica region (Hellenic Statistical Authority, 2024), is strongly influenced by anthropogenic activity.

~~The MAX-DOAS instrument is installed on a hill approximately 500 m above ground level to the east of the city (see Fig.2). Measurements are routinely conducted in multiple directions. For this analysis, we use data collected between January 2021~~

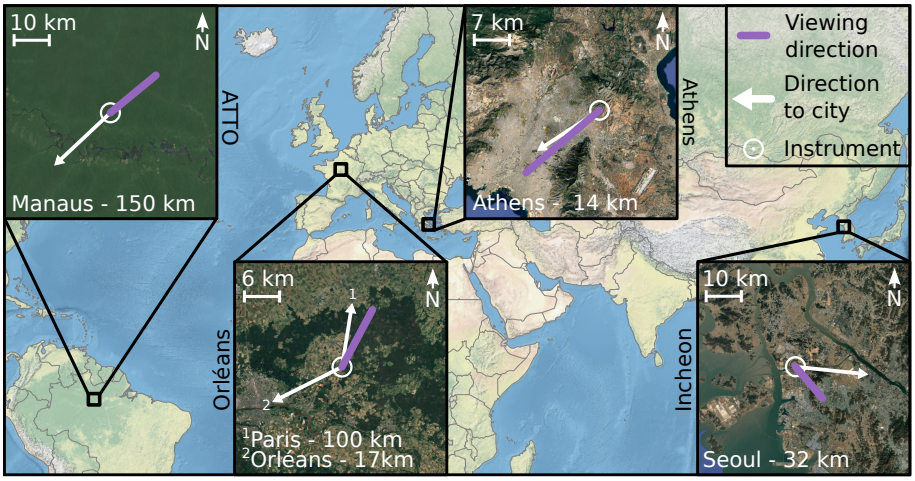

Figure 2. ~~World map~~ Map showing the location of all stations, their surroundings and distances to neighbouring cities. The white circles indicate the instrument positions, the white arrows show the direction to the ~~cities~~city centres, whereas the purple lines correspond to the relevant viewing direction of the instruments.

~~and December 2023 from the viewing direction oriented toward the city centre (232.5relative to north, indicated by the purple line in Fig. 2). Additional details regarding the instrument hardware and setup are given by Gratsea et al. (2016).~~

Under certain meteorological conditions, local topography causes pollutants to accumulate within the urban area (Kassomenos et al., 1995). Additionally, due to its hot and dry climate, Athens occasionally experiences wildfires, as observed, for example, in 2018 and 2024 (Lagouvardos et al., 2019; Castro-Melgar et al., 2025). Mountains with Mediterranean vegetation are located to the north. To the east, the landscape features mountainous vegetation interspersed with smaller residential areas, while to the south lie the airport and lower-density residential and industrial zones. The city centre of Athens and the port of Piraeus are situated to the southwest .

The MAX-DOAS instrument is installed on a building roof on a hill (500 m above sea level) to the east of the city (see Fig. 2). Measurements are routinely conducted in multiple directions. For this analysis, we use data collected between January 2021 and December 2023 from the viewing direction oriented toward the city centre (indicated by the purple line in Fig. 2). Additional details regarding the instrument hardware and setup are given by Gratsea et al. (2016). Meteorologically, the region experiences low precipitation, pronounced diurnal and seasonal cycles in short-wave radiation, and relatively high temperatures exhibiting clear seasonal and daily variations during our measurements (Fig. 3). The prevailing winds during the measurement period come from northern directions, frequently reaching speeds above 9 ms^{-1} (Fig. A4).

2.5.2 Orléans

The second station is located near Orléans, France, on the premises of a radio station in Traînou, which is regularly used for scientific measurements, including the ICOS project (Ramonet et al., 2025). Traînou (approx. 3500 inhabitants; Institut

national de la statistique et des études économiques, 2025) is situated about 100 km south of Paris and 17 km northeast of Orléans (116000 inhabitants; Institut national de la statistique et des études économiques, 2024). ~~The exact location and surroundings of the instrument are shown in Fig. 2.~~

Crucially for this study, the site is adjacent to a large forested region ~~.~~ (Fig. 2). The Orléans State Forest covers roughly 350 km^2, and comprises a mixture of broadleaf and evergreen tree species (Bello et al., 2019). Thus, this measurement site is ~~primarily~~ strongly influenced by biogenic activity, with minimal local anthropogenic emissions, although pollutant plumes from Paris can occasionally be detected under northerly winds.

The MAX-DOAS instrument is mounted on an elevated position (approx. 10 m above ground level), enabling low-elevation scans directly above the forest canopy. Data analysed in this study cover the period from July 2023 to July ~~2024~~2025, focusing on measurements taken towards the forest (~~28relative to north,~~ Fig. 2).

~~Meteorologically,~~ Orléans experiences strong seasonal and diurnal variations in short-wave radiation, though its maximum values are comparatively low~~er~~due to its higher latitude (Fig. 3). Precipitation is moderate without a clear seasonality. Temperatures are among the lowest of the investigated sites, with a less pronounced seasonal cycle than at Incheon and Athens. The prevailing wind direction is from the southwest, frequently exhibiting high wind speeds exceeding 9 ms^{-1} (Fig. A4).

2.5.3 Incheon

The third instrument was installed on the roof of the Environmental Satellite Center in Incheon, part of the Seoul Metropolitan Area (SMA) in South Korea. With approximately 3 million inhabitants, Incheon is South Korea's third-largest city. The SMA is the most densely populated region in the country (Kim et al., 2021).

~~The instrument's exact location and surroundings are shown in Fig. 2.~~ It is situated in an anthropogenically dominated environment, with Seoul city centre approximately 32 km to the east, Incheon city centre to the south, and the harbour area to the west ~~.~~ (Fig. 2). The northern edge of the metropolitan area borders North Korea (about 20 km north), where some forested mountains are located.

As part of the GEMS Map of Air Pollution (GMAP) 2021 campaign and the Satellite Integrated Joint Monitoring of Air Quality (SIJAQ) 2022 campaign, MAX-DOAS measurements were conducted for about one year. For this study, we analyse data from October 2021 to November 2022, focusing on the urban azimuth viewing direction~~at 137.5relative to north~~.

Meteorologically, ~~Incheon is strongly influenced by the East Asian monsoon. Heavy~~ heavy rainfall occurs between June and September, while the rest of the year is comparatively dry. This signal indicated the influence of the East Asian monsoon and tropical cyclones. No pronounced diurnal precipitation cycle is observed. The seasonal precipitation pattern affects short-wave radiation, which declines during the wet months but otherwise shows strong seasonal and diurnal cycles with high peak values. Temperatures also exhibit strong seasonal and diurnal variability, with the lowest temperatures across all sites recorded in December and January. The prevailing wind direction is from the northwest (especially during Winter) or west (Fig. A4).

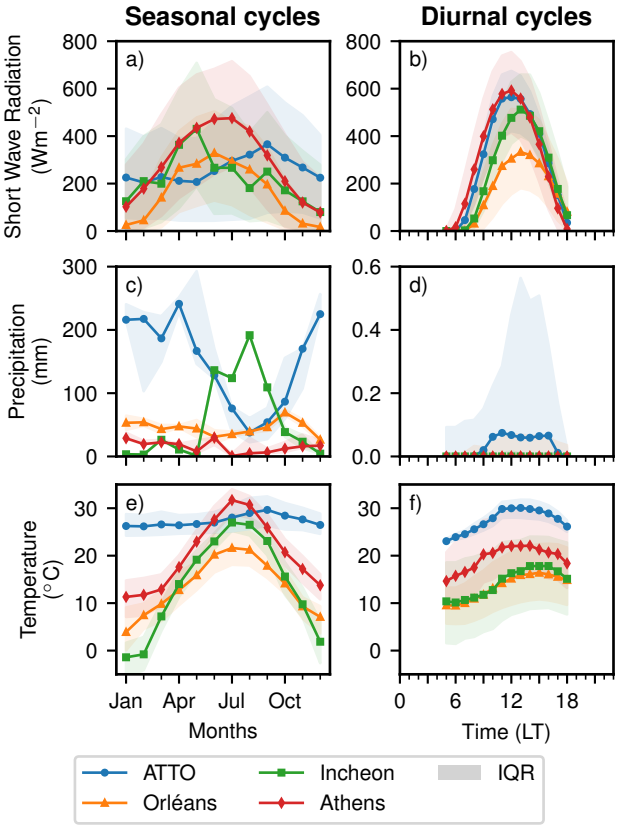

Figure 3. Meteorological overview showing the seasonal cycles (left column) and the diurnal cycles (right column) of ~~mean~~ median short wave radiation (first row), median monthly/hourly sum of precipitation (second row) and ~~mean~~ median temperature (third row) for the analysed stations based on ERA5 data~~for~~. The shading corresponds to the ~~individual measurement periods~~interquartile range (IQR). To closely describe the conditions during measurements, only data during daytime, between 5:00 and 18:00 local time, are considered during the sites operation years.

2.5.4 ATTO

The fourth ~~station~~ instrument is located on the tall ATTO Tower in Brazil, deep within the Amazon rainforest. Situated approximately 150 km northeast of Manaus (population 2 million, (Instituto Brasileiro de Geografia e Estatística, 2022)), the ATTO site serves as a remote research site in the heart of the rainforest (Fig. 2). The surrounding area is sparsely populated, resulting in the site being predominantly influenced by biogenic activity. During the dry season wildfires are more frequent and affect local atmospheric conditions (Andreae et al., 2015; Donner, 2024).

The instrument was installed at a height of 80 m in October 2017 and measurements are still ongoing at the time of writing. However, not all data are yet analysed in scientific quality, so the used dataset ends in August 2022. Some data gaps occurred due to the challenging hot and humid climate affecting hardware and electronics. The ~~instrument is oriented northeast (50°~~

Table 1. Station information overview.

	ATTO	Orléans	Athens	Incheon
instrument position	2.15 ° S, 59.00 ° W	47.96 ° N, 2.11 ° E	38.05 ° N, 23.86 ° E	37.57 ° N, 126.64 ° E
instrument type	multi-axis	multi-axis	multi-axis	multi-axis
station altitude (asl)	120 m	130 m	500 m	0 m
instrument altitude (agl)	80 m	10 m	5 m	20 m
viewing elevation[a]	2 °[b]	1 °	1 °	3 °[c]
azimuth viewing direction[d]	50 °	28 °	232.5 °	137.5 °
period start	12.10.2017	03.07.2023	01.01.2021	06.10.2021
period end	31.07.2022	01.07.2025	31.12.2023	15.11.2022
period coverage[e]	66 %	72 %	64 %	79 %
median NO_2 dSCD[f] (molec cm^{-2})	9.61×10^{14}	1.47×10^{16}	5.73×10^{16}	9.71×10^{16}
IQR NO_2 dSCD[f] (molec cm^{-2})	$(3.94–18.0) \times 10^{14}$	$(1.09–2.18) \times 10^{16}$	$(3.53–9.56) \times 10^{16}$	$(6.60–14.2) \times 10^{16}$
NOx/VOCs emission ratio[g]	0 %	26 %	50 %	66 %
AVOCs/BVOCs emission ratio[h]	0 %	174 %	2500 %	5800 %

[a] The supplement also contains figures with data from 30 ° viewing elevation.
[b] The highest O_4 dSCDs occurred for 2 ° elevation
[c] Lower elevations are obstructed
[d] N= 0 ° and E= 90 °
[e] Days with observations after filtering and merging (intersect) of all trace gases
[f] From viewing elevation
[g] Based on area weighted annual average CAMS-GLOB-ANT emissions and CAMS-GLOB-BIO emissions during measurement years (excluding 2025 for Orléans).
[h] Ratio of anthropogenic non-methane VOCs (AVOCs) emissions from CAMS-GLOB-ANT to biogenic non-methane VOCs (BVOCs) emissions from CAMS-GLOB-BIO.

~~relative to north). The~~ dataset analysed in this study was originally obtained by Donner (2024), who also ~~provided~~ provides a detailed description of the site and instrumentation. Some figures from that publication are reproduced here using our own processing methodology based on their dataset. In such cases, the figure captions indicate which panels are affected.

Meteorologically, the ATTO Tower is characterised by a tropical climate, see Fig. 3. Precipitation ~~predominantly occurs during~~ is largely confined to the wet season (December–May), with much drier conditions prevailing during the rest of the year. ~~A pronounced diurnal precipitation cycle is observed~~Within this season, rainfall typically occurs between 10:00 and 16:00 local time. Temperatures are consistently high, showing daily but minimal annual variation. Short-wave radiation exhibits a strong diurnal pattern but remains relatively stable on seasonal timescales, with only a slight reduction during the wet season. Prevailing winds are from the northeast, but compared to the other sites, wind speeds are predominantly low, typically below 3 ms^{-1} (Fig. A4).

2.5.5 Coverage and representativeness

The four stations cover a broad range of environmental conditions; however, they cannot represent the full diversity of atmospheric regimes. In particular, both urban sites, Incheon and Athens, are located near the coastline, implying potential influences from marine air masses and sea-salt aerosols that may not be representative of inland urban environments. The datasets were collected during non-overlapping periods, as the station locations originate from long-term measurement activities. While this limits strict temporal comparability, the analysis focuses on characteristic relationships rather than direct year-to-year contrasts.

The horizontal orientation of the light paths introduces an additional spatial averaging that is inherent to MAX-DOAS measurements and is illustrated in Fig. 2. The retrieved dSCDs represent the concentration along the effective light path, whose length within the boundary layer depends on atmospheric visibility. Under clear conditions, photons scattered at distances of up to approximately 15 km from the instrument can contribute to the signal (Seyler et al., 2017).

Beyond viewing geometry, the origin and transport history of observed air masses determine the spatial representativeness of each site. Annual and seasonal horizontal footprints derived from backward simulations show, as expected, the highest sensitivity in the vicinity of each instrument (Fig. 4). The ATTO footprint shifts seasonally with the movement of the ITCZ. At Orléans, persistent sensitivity to both the city and the surrounding forest reflects mixed anthropogenic and biogenic influences, with enhanced sensitivity towards Paris (100 km to the northeast) during MAM. Athens exhibits strong sensitivity to the urban area and more remote northern regions, with reduced sensitivity to the city centre and harbour in JJA. Incheon shows pronounced sensitivity to northwestern source regions during DJF, with a less directional footprint in the remaining seasons. Overall, the footprint analysis is consistent with the site descriptions and classifications, but reveals minor seasonal sampling biases that should be considered when comparing sites.

3 Results and discussion

~~In this chapter, we investigate multiple factors influencing R^*_{GF} to assess its potential as an indicator for VOC source identification. As outlined in the introduction, previous studies have identified a wide range of influencing factors. This work specifically examines diurnal, weekly, and seasonal cycles, as well as the dependence of R^*_{GF} on elevation angle, temperature, and NO_2 levels across all four stations.~~

3.1 ~~Diurnal cycle~~Temporal cycles

3.1.1 Diurnal cycle

A diurnal cycle describes the variation over a day. It allows to compare with other variables that change regularly over the day, e.g. incoming solar radiation or car traffic. For the case of R_{GF}, multiple diurnal cycles were reported. At two sites, one in India (semi-urban) and one in Thailand (rural), ~~Hoque et al. (2018a, b); Rawat et al. (2024)~~ Hoque et al. (2018a, b) and Rawat et al. (2024) observed a diurnal cycle with a noon maximum for R_{GF} based on VCDs retrieved from MAX-DOAS

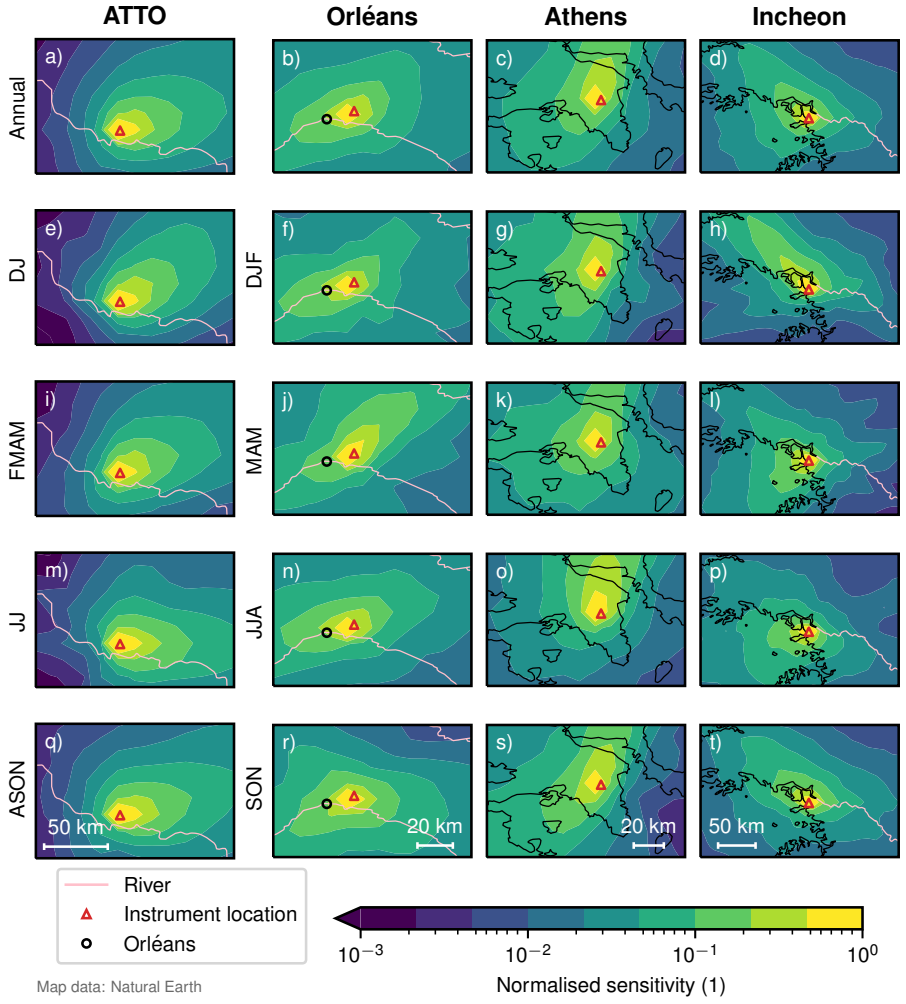

Figure 4. Annual and seasonal station footprints, based on normalised sensitivity with respect to the maximum per panel. The annual distribution is shown in the first row, and the seasonal distributions are shown in the rows below. Note that the months for ATTO are grouped differently to account for wet (FMAM) and dry (ASON) season.

measurements. The values ranged from 2–4 %. Hoque et al. (2018a) also found the diurnal cycle of R_{GF} to be less pronounced in the dry season compared to the wet season, which we will revisit for ATTO in Sect. ~~??~~3.1.1.

DiGangi et al. (2012) investigated the diurnal cycle at two predominantly biogenic sites at higher altitudes (Sierra Nevada Mountains, 1315 m; Rocky Mountains 2286 m) with field campaigns in July 2009 and August 2010 utilizing in-situ instruments. The average R_{GF} values for both campaigns were about 2 % and 1.7 %. At the Sierra Nevada site, ~~RGF~~ R_{GF} increased to about 3 % around midday to ~~to~~ afternoon. Whereas the Rocky Mountains campaign showed only minor diurnal variability.

DiGangi et al. (2012) attributed the observed enhancements primarily to anthropogenic VOCs and biomass burning plumes encountered during the campaigns.

The diurnal cycles of R^*_{GF} ~~for our dataset (all four stations) are shown in Fig. ??, plotted against~~ in local solar time (LST) ~~. The diurnal cycles~~ differ strongly across the ~~stations.~~ four stations (Fig. 5). Anthropogenic sites show pronounced diurnal variability, while the biogenic sites show relatively little variation over the day. At ATTO and Orléans, the diurnal cycles are relatively flat~~, with average values of 3.1±0.4 and 2.2±0.2 , respectively~~. In contrast, Athens and Incheon exhibit higher average values ~~(3.5±0.4 and 3.7±0.7) and~~ and strong diurnal patterns, with peaks around 10 ~~am~~ :00 in Athens and noon in Incheon. In Athens, the cycle follows morning rush hour, whereas Incheon has a noon maximum, indicating different drivers of R^*_{GF} over the day for both cities.

The diurnal pattern of R^*_{GF} is broadly consistent across seasons, although seasonal offsets in the absolute values are present. The largest offset occurs between the wet and dry seasons at ATTO, which is discussed in detail in Sect. 3.1.1. Notably, although the diurnal cycles of HCHO and CHOCHO individually change between seasons, their ratio R^*_{GF} retains a similar diurnal shape throughout the year (Fig. A1).

Except at ATTO, data availability decreases in the early morning and late afternoon due to the applied SZA filter, which excludes measurements ~~when the sun is low.~~at large SZA. 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. Data from these times of day are mainly collected during summer months, introducing a seasonal bias in the early and late portion of the diurnal cycle. ~~Despite this, the diurnal patterns are representative across the year, although seasonal offsets are present. It is important to note that the~~ Furthermore, the number of valid data points decreases substantially during the winter months at Orléans, Incheon, and Athens, primarily because we filter by relative error of CHOCHO dSCDs, which increases for low atmospheric concentrations. Filtering by relative error is needed to limit the scatter of R^*_{GF}, but it means that R^*_{GF} is more representative for high CHOCHO and HCHO columns.

The effect of scaling the vis O_4 dSCDs for ATTO is highlighted again by showing the original data with the blue lines in Fig. ??+??5. The overall high O_4 dSCDs in the visible lead to a really low O_4 ratio (Fig. ??i5m), which is mirrored in the overall low level of R^*_{GF} (Fig. ??5a).

~~Examining the three components separately (dSCDs in the top row, dSCDs in the center row, and the ratio in the bottom row of Fig. ??) reveals that behaves differently across the four stations.~~ The O_4 ratio, after scaling ATTO, is of similar magnitude across sites and does not contribute to a pronounced diurnal cycle.

Examining the components of R^*_{GF} in Fig. 5 reveals that CHOCHO behaves differently across the four stations. HCHO dSCDs follow a U-shaped diurnal cycle at all stations, with a maximum in the morning and evening and a minimum around noon. This pattern has previously been attributed to enhanced sinks (photolysis and OH oxidation) dominating around midday (Nussbaumer et al., 2021; Donner, 2024). However, the underlying processes are more complex as they can also promote secondary formation of CHOCHO and HCHO by breaking down VOC precursors. ~~In addition, potential dilution effects associated with boundary layer growth may contribute to the observed shape.~~

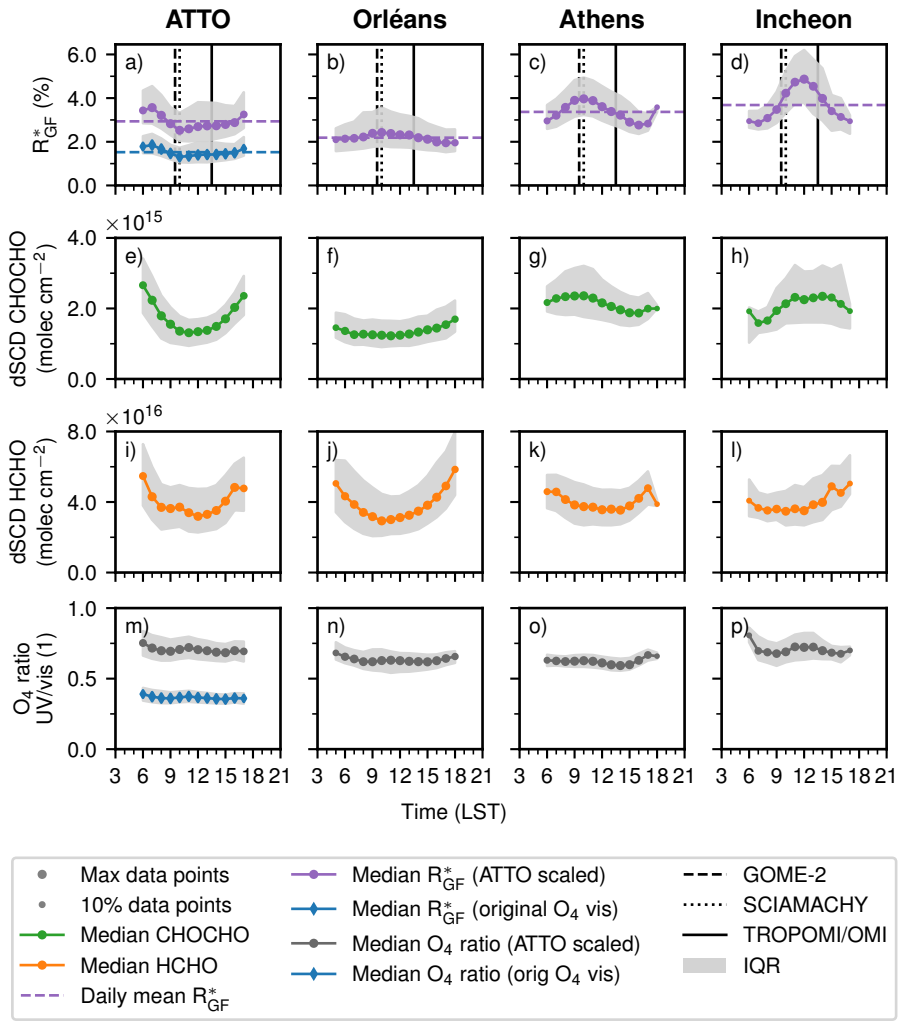

Figure 5. Diurnal cycles of R^*_{GF} (top row), CHOCHO dSCD (upper centre row), HCHO dSCD (lower centre row), and O_4 ratio (bottom row) for ATTO, Orléans, Athens, and Incheon relative to local solar time (LST). Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements. The original R^*_{GF} and O_4 ratio without scaling O_4 vis dSCDs is shown for ATTO in blue with diamond markers. In addition, the overpass times of GOME-2, SCIAMACHY and TROPOMI/OMI are highlighted with black vertical bars. Panels e) and i) are self-created based on Donner (2024).

The diurnal cycle of the CHOCHO dSCDs varies in magnitude and shape across stations. ATTO also shows a pronounced U-shape. Orléans exhibits a relatively flat diurnal cycle. Contrasting to that, the anthropogenic stations show a different behaviour. Here, we find higher daily averages plus a maximum in Athens around 10~~am~~:00 and in Incheon over noon. The ~~diurnal cycle shapes of~~ shapes of the diurnal cycles at the anthropogenic stations suggest a stronger link to anthropogenic activity ~~, and their difference from~~ for CHOCHO than HCHO~~indicates a weaker anthropogenic influence on~~. Since direct CHOCHO

emissions are suspected to be low (Stavrakou et al., 2009a; Silva et al., 2018), anthropogenically emitted precursors with a high CHOCHO yield might be a possible explanation, like aromatics (Chan Miller et al., 2016) or acetylene/ethylene (Fu et al., 2008). Furthermore, other effects independent from emissions could have an influence, like differences in photolysis, OH loss, heterogeneous uptake, and wet removal, but our dataset does not allow to separate such effects. Resulting different photochemical lifetimes of CHOCHO and HCHO might also contribute to the shape of the diurnal cycle of R^*_{GF}. Under simplified conditions, a longer lifetime of CHOCHO compared to HCHO, results in an increase of R^*_{GF} and a decrease otherwise (see Fig. A11).

Considering these curves, the diurnal cycle of R^*_{GF} appears to be driven by CHOCHO. The enhanced daily mean R^*_{GF} in Athens and Incheon can be explained by the overall higher CHOCHO levels. The shape of R^*_{GF} diurnal cycle can be attributed to the behaviour of CHOCHO dSCDs. Similar shapes between CHOCHO dSCDs and HCHO dSCDs lead to flat cycles at ATTO and Orléans, whereas the different shapes of CHOCHO dSCDs and HCHO dSCDs lead to a pronounced diurnal cycle of R^*_{GF} at the anthropogenic stations.

~~Compared~~ A direct quantitative comparison with previous studies ~~, these observations are consistent with the reported midday peaks in R_{GF} at sites in Thailand (described as rural) and India (described as semi-urban) (Hoque et al., 2018a, b; Rawat et al., 202~~ , 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 ~~moderate to flat diurnal cycles with lower absolute RGF (<2) at predominantly biogenic sites(DiGangi et al., 2012)~~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.

In summary, R^*_{GF} shows enhanced average values over the day for anthropogenic stations, due to enhanced CHOCHO levels. This indicates that R^*_{GF} contains information about the different environments, which supports its usage as a proxy for VOC origin. The pronounced diurnal cycles for anthropogenic stations, however, complicate the interpretation as the timing of the measurement becomes important. The implications for comparing R_{GF} values of different studies are discussed in Sect. 3.4.3.

3.2 ~~Seasonal cycle~~

3.1.1 Seasonal cycle

The variation over the year, the seasonal cycle, enables to investigate how ~~the~~ a variable is connected to changes of other variables based on seasons. Multiple studies have reported seasonal cycles for R_{GF} so far: ~~Hoque et al. (2018a, b); Rawat et al. (2024)~~ Hoque et al. (2018a, b) and Rawat et al. (2024) found a relatively flat seasonal pattern at Pantnagar (India, described as semi-

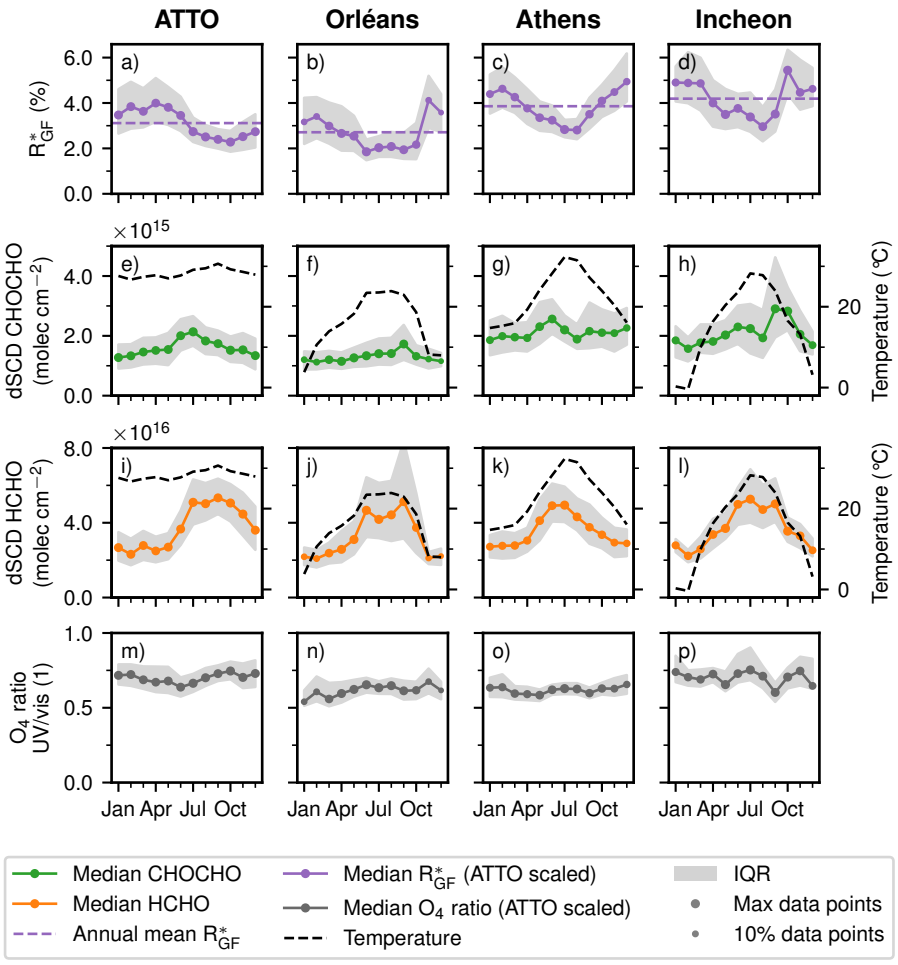

Figure 6. Seasonal cycle of R^*_{GF} (top row), CHOCHO dSCD (upper centre row), HCHO dSCD (lower centre row), and O_4 ratio (bottom row) for ATTO, Orléans, Athens, and Incheon. Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements. The seasonal cycle of temperature is shown on a secondary axis with a dashed black line. Panels e) and i) are are self-created based on Donner (2024).

urban) based on MAX-DOAS VCDs. At a second site, Phimai (Thailand, described as rural), the seasonal cycle showed an increase from January to September. Similarly, Xing et al. (2025), analysing one year of MAX-DOAS VCDs from Guangzhou (China), found enhanced R_{GF} values from November to April and lower values during the rest of the year.

~~The seasonal cycles of R^*_{GF} for our dataset of all four stations are shown in Fig. ??. The annual mean R^*_{GF} of the stations ranges from 2.8±0.7 in Orléans to 4.2±0.8 at Incheon.~~

The overall shape of the seasonal cycle of R^*_{GF} is similar across all four stations, with one minimum and one maximum per year ~~.~~ (Fig. 6). At Orléans, Athens, and Incheon, the lowest values occur in July and August (late summer), while the highest

values are observed between October and March (winter). At ATTO, the seasonal cycle is shifted by several months, with a minimum in October (dry season) and a maximum extending into June (wet season). Notably, the minimum phase at the biogenic sites tends to be more prolonged compared to the anthropogenic sites. Fewer data points are available in winter due to filtering based on relative error (see Sect. 3.1.1).

~~Examining the three components separately (dSCDs in the top row, dSCDs in the center row, and the ratio in the bottom row of Fig. ??)~~ Examining the components of R^*_{GF} separately reveals that both trace gases behave differently for all four stations, whereas the O_4 ratio is similar. Looking at the seasonal cycles of HCHO dSCDs we see one enhanced period during the middle of the year, a narrow peak during June and July for Athens, and an extended peak over four months spanning from June to October for the other stations. The annual means and the amplitude are comparable between the stations.

The seasonal cycle of CHOCHO dSCDs is relatively flat with one peak in different months from June (Athens) to October (Incheon). One can see a shift to higher annual mean values from ATTO to Incheon. The anthropogenic stations show the highest CHOCHO dSCDs and more variability over the year.

~~The second axis in Fig. ?? shows the temperature at 2 from ERA5, matched to each measurement and then binned in conjunction with the dSCDs.~~ 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 ($t = -5.8, p = 8 \times 10^{-8}$) between the biogenic and anthropogenic group. A Welch-ANOVA ($F = 19, p = 3 \times 10^{-8}$) 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.

Three seasonal shifts of the station footprints were identified from Fig. 4 for the non-tropical sites: increased sensitivity toward Paris during MAM at Orléans, reduced sensitivity to the harbour and city centre during JJA at Athens, and enhanced sensitivity to less densely populated regions northwest of Incheon during DJF. None of these shifts is clearly reflected in the seasonal cycle of R^*_{GF}, suggesting that a simple seasonal categorisation might not be enough to capture clear pathway dependencies (Poulidis and Takemi, 2016).

For all stations except ATTO, the seasonal cycle of HCHO dSCDs closely resembles the seasonal cycle of temperature~~.~~ , which is shown on a second axis in Fig. 6. While temperature variability in ATTO is limited, a slight increase during September–October coincides with peak HCHO values. The seasonal cycle of CHOCHO does not show a clear influence by temperature. This ~~comparison~~ highlights ATTOs unique tropical conditions: ~~The temperature does not change significantly over the year and therefore~~ the near-constant temperature throughout the year means that seasonal variability in both trace gases ~~are driven by different processes~~ is governed by processes other than temperature.

Donner (2024) suggested, that ~~both~~ the two trace gases undergo different processing in the dry and wet season and that ~~both~~ the seasons probably have a different precursor composition. ~~Figure 7 shows the diurnal cycle of R^*_{GF} during the wet and dry season.~~ As reported by Donner (2024), enhanced R^*_{GF} values in the wet season and reduced values in the dry season are found~~.~~

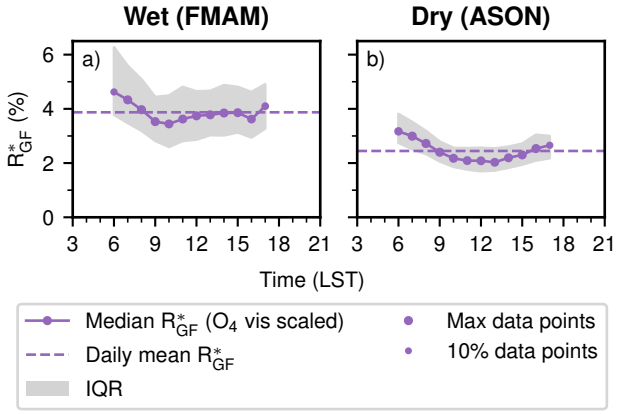

Figure 7. Diurnal cycles in the wet a) and dry b) season of R^*_{GF} at ATTO. ~~To indicate~~ Marker size scales with the number of ~~measurements that contribute to each data point~~contributing observations, ~~the size of the marker is varied~~with smaller markers indicating fewer measurements.~~The solid line corresponds to the median value, the dashed line to the daily mean, and the gray shaded area to the IQR. Only data from the lowest viewing elevation is included.~~

, see Fig. 7. The daily mean is reduced by 0.7~~%pt.~~ %pt. in the dry season. The shape of the diurnal cycle is relatively flat. Since forest fires predominantly occur in the dry period, previously excluded pyrogenic activity may contribute to the observed changes. This would be supported by enhanced NO_2 levels and aerosols during the dry season as shown by Donner (2024). ~~However as~~ Since biomass burning has been reported to lead to higher R_{GF} levels (DiGangi et al., 2012; Zarzana et al., 2017; Chan Miller et al., 2014; Alvarado et al., 2020), the ~~lower R_{GF} levels~~ observed low median values at this site are unlikely to ~~originate from pyrogenic activity in the vicinity~~reflect a significant pyrogenic contribution. Individual pyrogenic events may nonetheless produce enhancements in R^*_{GF} that are not captured in the median.

The discrepancy between wet and dry season is in agreement with the findings of Hoque et al. (2018a), where they found higher R_{GF} during the wet season and lower R_{GF} during the dry season in Phimai (Thailand). Furthermore, both seasons share the same diurnal cycle for Hoque et al. (2018a). However, their diurnal cycle had a pronounced noon maximum, which is not present in this dataset, ~~which~~ and might, even though the Phimai site is described as rural, hint at a stronger anthropogenic influence than at ATTO, see Sect. 3.1.1.

Having these points in mind, the seasonal cycle of R^*_{GF} seems to be driven, contrary to the diurnal cycles, by the variability of HCHO. The variability of HCHO dSCDs is strongly connected to the variability of temperature for non-tropical stations and seems to be connected to the dry/wet season for ATTO. The enhanced annual mean R^*_{GF} at anthropogenic stations can be explained by overall higher CHOCHO levels.

~~Comparing with the previous studies, these observations of anthropogenic stations match best with the enhancements during the colder months~~ 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) ~~. Comparing the absolute values, lower absolute~~ for Guangzhou. Our absolute R^*_{GF} values are ~~observed, but caution is required, as the~~ 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 ~~based on corrected dSCDs for the lowest viewing elevation instead of VCDs. Contrary to Hoque et al. (2018b), no increase over the year is visible in our dataset, but a similar magnitude for remote stations is found in this study.~~ comparable to that reported by Hoque et al. (2018b) for Pantnagar, even though no progressive annual increase is observed like at Phimai.

Chen et al. (2023) published global R_{GF} maps based on the TROPOMI observations for the year 2019. Although our R^*_{GF} is derived from dSCDs and therefore does not correspond to the exact same measurement volume (Sect. 3.4.1), a comparison of the magnitude of annual means is still meaningful. Extracting R_{GF} values at our measurement sites from their maps for 2019 suggests the following ranking: Incheon > ATTO > Orléans. Athens could not be identified in their maps due to its vicinity to the coastline. Furthermore, Chen et al. (2023) maps show enhanced R_{GF} values during the wet season compared to the dry season at the ATTO site, which is consistent with our observations.

To conclude, we see a similar pattern for seasonal cycles as for diurnal cycles: R^*_{GF} exhibits a cycle and its average and its amplitude are more pronounced for anthropogenic stations, but this time originating from variations in HCHO. This complicates the interpretation of R^*_{GF} as a proxy for VOC origin, because many other seasonal effects can contribute to its variation, e.g. temperature, which are difficult to disentangle from changes in VOC origin over the year. Moreover, longer time series are needed for measurement campaigns to avoid sampling biases.

3.2 ~~Weekly cycle~~

3.1.1 Weekly cycle

As anthropogenic emissions are typically lower on the weekend, the weekly cycles can be used as an indicator for the contribution of anthropogenic emissions (Beirle et al., 2003). Gratsea et al. (2016) reported a weekly cycle in Athens for glyoxal and to a lesser extent for formaldehyde, but only for measurements dominated by urban air.

To our knowledge, no previous study has investigated weekly cycles specifically for R_{GF}, but from the findings of Gratsea et al. (2016), we expect a weekend effect may occur. ~~The weekly cycle of R^*_{GF} for our four sites is shown in~~ In Fig. **??**. 8, ATTO and Orléans display flat weekly cycles, while the anthropogenic stations show an offset between weekday and weekend. Moreover, as seen and discussed before for diurnal and seasonal cycles, the average value throughout the week is higher for the anthropogenic stations. It should be noted, that R^*_{GF} for Incheon is reduced not only during the weekend but also on mondays.

~~Based on the four sites, we observe a weekend effect in~~ For both stations, the mean weekday R^*_{GF} ~~at the anthropogenic sites, Incheon and Athens. Specifically, R^*_{GF} decreases on weekends in both Incheon and Athens, with the strongest reduction seen in Athens by up to 0.4~~exceeds the mean weekend value by 0.5 %pt., ~~i.e. about~~ corresponding to a reduction of approximately 10 % ~~lower than during weekdays.~~ on weekends. Although this relative difference is comparable to our systematic uncertainties, these uncertainties are expected to affect all days uniformly and should therefore not be relevant for the weekend effect. As

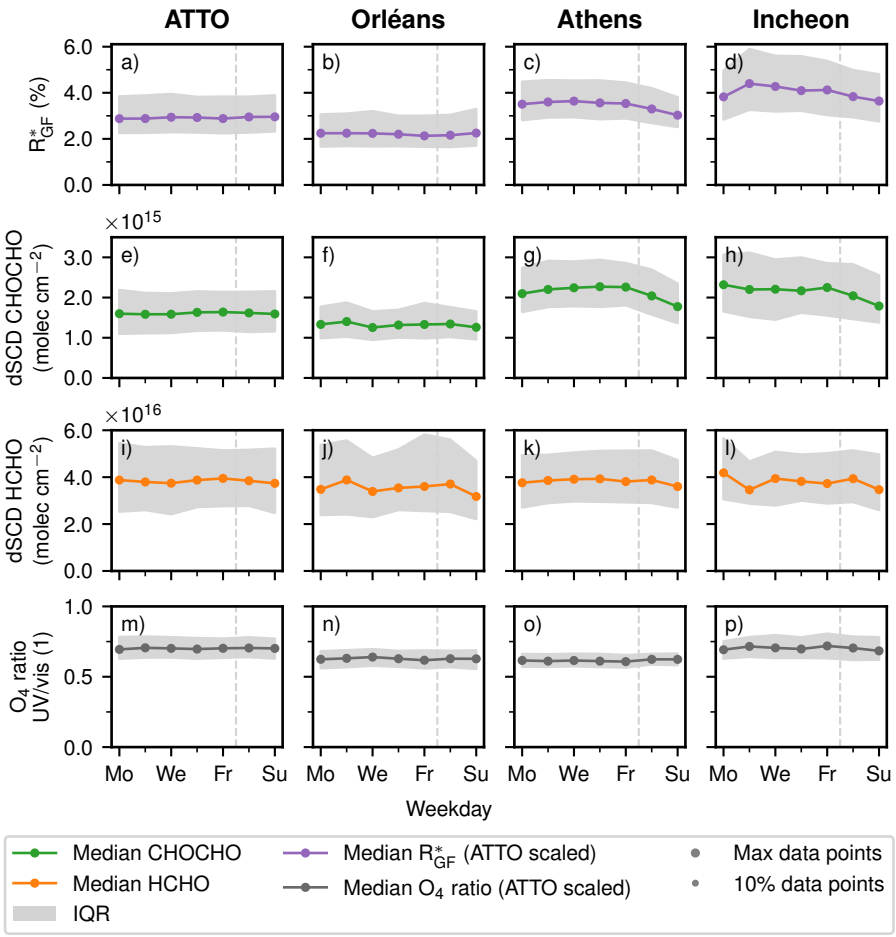

Figure 8. Weekly cycle of R^*_{GF} (top row), CHOCHO dSCD (upper centre row), HCHO dSCD (lower centre row), and O_4 ratio (bottom row) for ATTO, Orléans, Athens, and Incheon. Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements.

described in Sect. 2.4, the differences between weekday and weekend are significant for Athens ($t = 4.4$, $p = 2 \times 10^{-5}$) and Incheon ($t = 2.7$, $p = 8 \times 10^{-3}$).

~~In Figure ?? weekly cycles of both oxygenated VOCs (OVOCs) and the ratio are shown.~~

For ATTO and Orléans the weekly cycles of both OVOCs are relatively flat and show no weekend effect ~~.~~ (Fig. 8). Comparing both OVOCs over the week for the anthropogenic stations, CHOCHO dSCDs show a strong weekend effect for Athens and Incheon. HCHO dSCDs, however, do not show a strong decrease on the weekend, therefore the weekend effect observed for R^*_{GF} is driven by the weekend effect from CHOCHO dSCDs.

To summarize, R^*_{GF} exhibits a weekend effect for anthropogenic stations, driven by the weekend effect of CHOCHO dSCDs. Showing a weekend effect supports R^*_{GF} usage as a proxy for different VOC origin, as changes in anthropogenic emissions are mirrored in R^*_{GF}.

3.2 ~~Elevation dependence~~Link to meteorology

~~Investigating the elevation dependence of~~ Atmospheric levels of VOCs are known to be influenced by temperature (Pusede et al., 2014; Bou
, which could also impact R_{GF}. In addition to temperature, several meteorological factors could theoretically affect R_{GF}. For example, enhanced photolysis rates under higher short wave radiation may alter production and loss pathways, while increased aerosol liquid water content could enhance aerosol uptake and wet deposition of CHOCHO. We use ERA5 meteorological data to examine the dependence of R_{GF} ~~provides insight into how R_{GF} is associated with different altitudes. Previous studies that examined the altitude dependence of R_{GF} reported varying results. Kaiser et al. (2015) observed a slight increase in R_{GF} in the free troposphere compared to the boundary layer. Similarly, Xing et al. (2020) reported lower values within the lowest 100 and higher values up to 1 altitude. In contrast, Rawat et al. (2024) found relatively stable R_{GF} values up to approximately 3.7 , followed by a sharp decrease above that level. Hong et al. (2024), studying Chinese cities, reported a strong increase in R_{GF} with altitude, up to 2 .~~ on temperature, dew point temperature, relative humidity, boundary-layer height, short wave radiation, and wind speed.

~~As we use dSCDs for~~ Although R^*_{GF} ~~in our analysis, we can only investigate elevation dependence , rather than true altitude dependence. However, the elevation dependence still hints at the altitude distribution in the way that the signal from different elevation angles is dominated by different altitudes.~~

~~R^*_{GF} (with correction) and R_{GF} (without correction) at different viewing elevations for ATTO, Orléans, Athens, and Incheon. The stations are sorted from left to right from the most biogenic environment (ATTO) to the most anthropogenic environment (Incheon). The line within the coloured box indicates the median value, the box itself the IQR. The whiskers extend to 1.5 IQR. The vertical dashed light gray line indicates the transition to higher viewing elevations. To indicate the number of measurements that contribute to each box, the transparency of the box is varied (lower transparency meaning lower number of measurements). Note, that the lowest viewing elevation differs across the stations.~~

~~Since the correction is motivated by the lowest elevations close to the surface, Fig. **??** shows R^*_{GF} and the uncorrected R_{GF}. Interpretation of R^*_{GF} at higher elevation angles requires caution, which is why both quantities are shown. R^*_{GF} increases towards higher elevation angles at all four sites. At lower elevation angles R^*_{GF} is relatively constant.~~

~~Looking at the three components, see~~ exhibits variability with each of these parameters (Fig. ~~**??**, we see that and dSCDs behave consistently across the stations, but differently when directly compared. dSCDs drop consistently towards higher elevation angles, whereas dSCDs stay constant for the lowest 2-3 elevation angles and drop afterwards. The enhanced levels at higher elevation indicate a different profile and thus a higher lifetime for compared to . Moreover,~~ A6), the ~~ratio shows an increase with elevation and approaches 1 for 30elevation, implying that no correction is applied at this angle.~~

~~dSCD (top row), dSCD (center row), and ratio (bottom row) at different viewing elevations for ATTO, Orléans, Athens, and Incheon. The line within the coloured box indicates the median value, the box itself the IQR. The whiskers extend to~~

~~1.5 IQR. The vertical dashed light gray line indicates the transition to higher viewing elevations. To indicate the number of~~
605 ~~measurements that contribute to each box, the transparency of the box is varied (lower transparency meaning lower number of measurements). Note, that the lowest viewing elevation differs among the stations.~~

~~The elevation dependency of~~ meteorological variables are strongly intercorrelated, preventing a clear attribution within our dataset. Further, analysing the median R^*_{GF} for different bins of temperature and moisture content (represented by dew point temperature) shows primarily variation of R^*_{GF} ~~is influenced by two effects. (1) the difference elevation behaviour of versus~~
610 ~~described above, (2)the elevation dependence of the correction. The light path differences in the UV and vis are strongest for the lowest elevations and disappear for 30direction. This behaviour amplifies the strong decrease of at higher elevations, leading to higher~~ with temperature (Fig. A7). Given the range of processes directly linked to temperature, such as biogenic emissions and temperature-dependent secondary formation rates, we expect the temperature to be the dominant contributor to the variability of R^*_{GF} ~~values at higher elevations.~~

615 ~~To summarize, it is unclear how R^*_{GF} changes with altitude. Our analysis, limited to different elevation angles, suggests that R^*_{GF} increases with elevation~~.

3.3 ~~Temperature dependence~~

3.2.1 Temperature dependence

Few studies have investigated meteorological influence on R_{GF} ~~. Multiple effects can contribute to the temperature dependence~~
620 ~~of the trace gases, e.g. biogenic emissions and secondary formation rates. A previous study from Guo et al. (2021)used~~ until time of writing, Guo et al. (2021), who analysed long-path DOAS ~~in Shanghai (China) during summer. They touched the temperature dependence~~ measurements in Shanghai during summer, mentioned an increase of R_{GF} ~~for their campaign duration and concluded that R_{GF} increases with temperature .~~

with temperature over their campaign period. The temperature dependence ~~for our stations is shown in Fig. **??**. Across all~~
625 ~~stations,~~ of R^*_{GF} exhibits a similar pattern across all stations (Fig. 9): at lower temperatures, values remain relatively stable with some fluctuations. However, starting from about 15 °C, R^*_{GF} decreases ~~hyperbolically,~~ with a maximum reduction of up to 1.9 %pt. observed at Athens.

~~The temperature dependence of the three components is shown in Fig. **??**. The~~ The O_4 ratio does not vary with temperature. Looking at HCHO dSCDs ~~(center row)~~ it is visible, that the HCHO levels grow exponentially with increasing temperatures
630 across all stations. ~~Comparing with~~ CHOCHO dSCDs ~~, we see that the dSCDs~~ also rise with temperature~~for ATTO and~~ , most clearly visible for Orléans ~~, but~~ and way less pronounced ~~. For Athensand Incheon the bulk of the dSCDs does not increase with temperature. However, the maximum values occur at the highest temperatures for both sites.~~ for ATTO, Athens, and Incheon.

~~The~~ An exponential behaviour is expected, especially for the biogenic stations, as biogenic emissions of precursors are known to increase exponentially with temperature (Guenther et al., 1993, 2012; Bourtsoukidis et al., 2024); higher temperatures
635 enhance biogenic activity, which in turn leads to greater VOC emissions. In addition, the secondary formation via OH oxidation should increase with temperature as reaction rates rise (Berg et al., 2024).

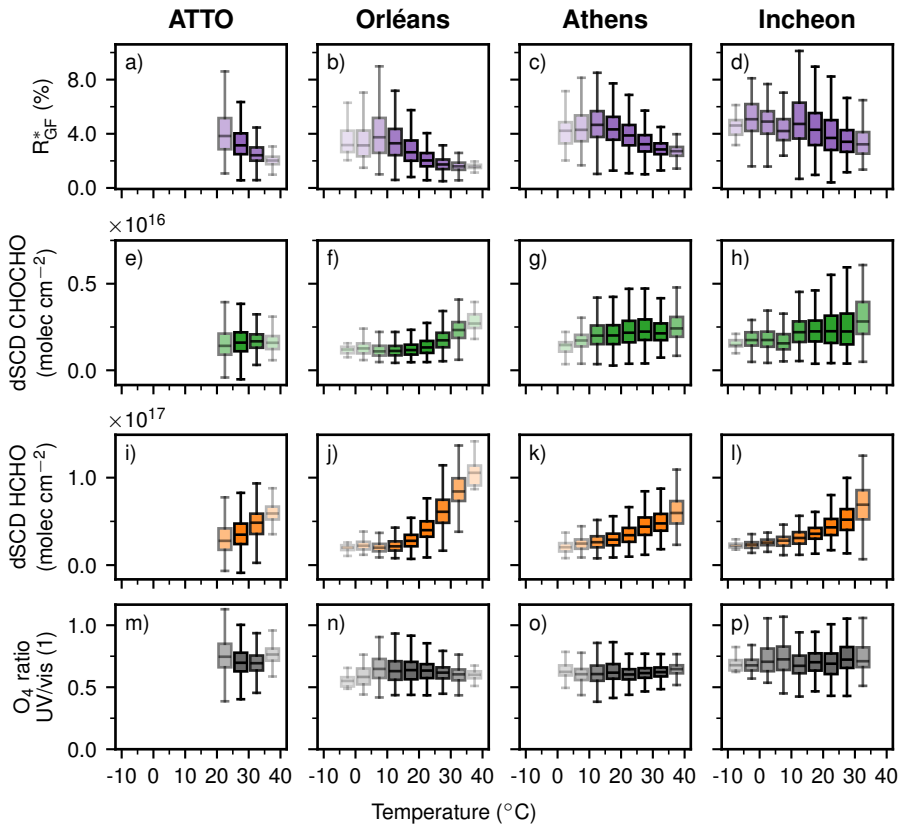

Figure 9. R^*_{GF} ~~at different temperatures intervals for ATTO a~~(top row), ~~Orléans b~~CHOCHO dSCD (upper centre row), ~~Incheon c~~HCHO dSCD (lower centre row), and ~~Athens d~~O_4 ratio (bottom row) ~~. The stations are sorted in as~~ a ~~clockwise order from the most biogenic environment (~~function of binned temperature for ATTO, ~~top left) to the most anthropogenic environment (Incheon~~Orléans, ~~bottom left)~~Athens, and Incheon. ~~The line within the coloured~~ Within each box, the horizontal line indicates the median ~~value,~~ and the box ~~itself~~ spans the IQR~~. The~~ ; whiskers extend to 1.5 IQR. ~~To indicate the number of measurements that contribute to each box, the~~ Box transparency ~~of~~ scales with the ~~box is varied (lower transparency meaning lower~~ number of contributing measurements~~)~~, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval. ~~Only data from the lowest viewing elevation is included~~Panels e) and i) are self-created based on Donner (2024).

~~Atmospheric levels of VOCs are known to be influenced by temperature (Pusede et al., 2014; Bourtsoukidis et al., 2024; Li et al., 2024), which could also impact~~

For the anthropogenic stations, ~~it is a differentsituation~~the situation is different. Here, we expect the anthropogenic emissions to be temperature independent and attribute the exponential increase with temperature primarily to the increased secondary formation at high temperatures. Adding to that, recent studies suggest that local biogenic VOC emissions in urban environments may play a more important role in local atmospheric chemistry than previously assumed (Liaskoni et al., 2024; Wang et al., 2025).

It is noteworthy that both trace gases do not behave identically at the anthropogenic stations. All above named arguments, increased secondary formation or potential local biogenic sources, hold for HCHO and CHOCHO, therefore, an important piece of information is still missing.

~~Overall, the temperature dependency of R^*_{GF} originates from the strong exponential increase of dSCDs across all stations, which starts around 15C. As dSCDs are used in the denominator, R^*_{GF} decreases hyperbolically.~~

To summarize, R^*_{GF} decreases with higher temperatures in our dataset, which is driven by the strong exponential increase with temperature ~~from~~ of HCHO. Similar to the other ~~chapters~~sections, this complicates the interpretation of R_{GF} as a proxy for VOC origin, as the R^*_{GF} values depend on temperature regardless of the environment of the sites. This has to be considered in the interpretation of R^*_{GF} when using simple thresholds.

3.2.2 Accounting for temperature dependence

As shown in the previous section, R^*_{GF} exhibits a strong dependence on temperature in our dataset. To isolate the variability not associated with temperature, we apply a regression-based correction to remove the temperature-correlated component. Specifically, deviations of R^*_{GF} from its arithmetic mean are fitted using an outlier-robust orthogonal linear regression. Influence from residuals exceeding two standard deviations from zero is reduced by applying linear weighting beyond this threshold. The fitted temperature-dependent component is then subtracted from the dataset.

After removal of the fitted component, the temperature-normalised R^*_{GF} (Fig. 10) shows only weak dependence on temperature. The diurnal cycles remain largely unchanged by the temperature correction, apart from a slightly reduced amplitude. In contrast, the seasonal variability is substantially reduced at all non-tropical sites, where the seasonal cycle nearly vanishes. At ATTO, however, the seasonal cycle persists, consistent with the comparatively small seasonal temperature variability in the tropics. Revisiting the remaining meteorological variables using the temperature-normalised R^*_{GF} shows that most of the previously observed variability disappears (see Fig. A2). This is consistent with the strong intercorrelations among the meteorological variables, indicating that the temperature-driven component can account for most of the variability seen for the other parameters.

Overall, removing the temperature-driven component largely eliminates both the apparent dependence of R^*_{GF} on other meteorological variables and the seasonal variability at the non-tropical sites. This suggests that temperature, or processes closely coupled to it, accounts for most of the observed seasonal variability in R^*_{GF}, while playing a lesser role in driving diurnal variability.

3.3 R_{GF}–NO_2 ~~dependence~~relationship

~~For~~ To assess the use of R_{GF} ~~usage to discriminate between different VOC origin~~for VOC source discrimination, it is important to examine how R_{GF} ~~reacts to changing~~ responds to changes in NO_2 levels, as NO_2 ~~is~~ serves as a good indicator ~~for~~ of anthropogenic activity.

~~Various~~ Several studies investigated the NO_2 dependency in the past with different results. Vrekoussis et al. (2010) using GOME-2 satellite data reported a clear link between R_{GF} and NO_2 levels, with ~~smaller~~ lower R_{GF} values found in polluted

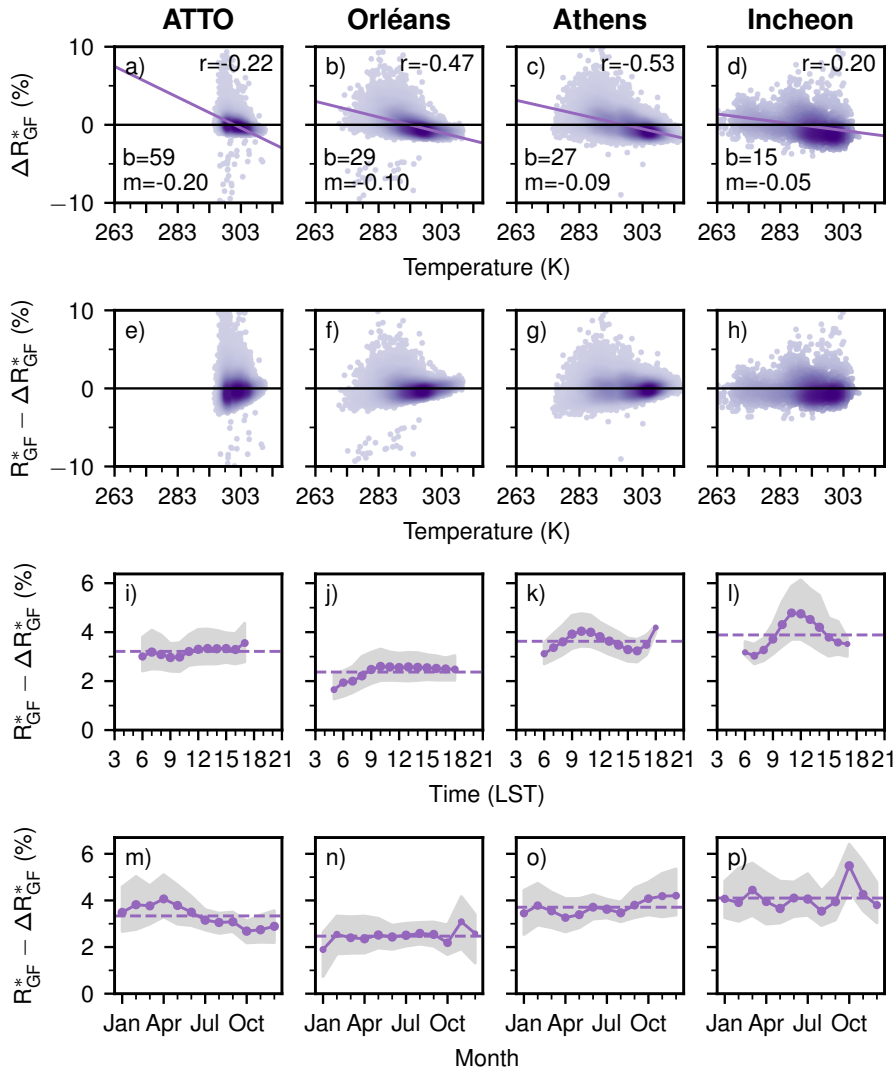

Figure 10. Panels a) to d) show deviation from the arithmetic mean R^*_{GF} and the respective regressions. Panels e) to h) show the temperature-normalised R^*_{GF} as a function of temperature. Panels i) to l) show the diurnal cycle and panels m) to p) the seasonal cycle for the temperature-normalised R^*_{GF}.

675 environments. Other studies ~~agreed on~~ supported this finding, e.g. Hoque et al. (2018a) using VCDs from MAX-DOAS in Phimai (Thailand), Xing et al. (2020) using VCDs from MAX-DOAS in Chongqing (China), ~~Hong et al. (2024) using with~~ and Hong et al. (2024) using VCDs from MAX-DOAS in four megacities (China). Chan Miller et al. (2017), however, observed no clear dependence on NO_2 levels using in-situ data from the flight-days with the SENEX aircraft. Another study, by Chen et al. (2023), using TROPOMI satellite data~~found~~, even reported the opposite trend, where R_{GF} increased with NO_2 levels.

Figure ?? shows the R^*_{GF} dependency on levels for all of our sites. The four stations cover Looking at Fig. 11, the four stations span a wide range of NO_2 dSCD values, from 10^{14} to 10^{18} molecules cm^{-2}. R^*_{GF} does not show a common behaviour for all stations . They can be consistent behaviour across all stations and can be broadly grouped in two categories: stations where no clear correlation between R^*_{GF} and NO_2 is observed (ATTO and Incheon) and stations where R^*_{GF} increases with higher NO_2 dSCD (Orléans and Athens).

Figure ?? presents how both OVOCs and the ratio depend on Both HCHO and CHOCHO increase with NO_2 levels. For for the first group of stations where (ATTO and Incheon), and this effect cancels out in the R^*_{GF} shows no clear correlation to , we see that both trace gases increase with and this is cancelled in the ratio. For the other second group, we see a different behaviour in each station. In Orléans, HCHO dSCDs decrease with higher NO_2 levels and therefore , and therefore, R^*_{GF} increases. In Athens, CHOCHO dSCDs increasesmore rapidly with higher NO_2 levels compared to HCHO dSCDs . This rapid increase is mirrored in the increased resulting in increasing R^*_{GF} at high NO_2 levels.

The different behaviour of both anthropogenic stationsis interesting. The differing behaviour of the two anthropogenic stations, Athens and Incheon, is noteworthy. Despite both being urban environments, R^*_{GF} exhibits a different behaviour even though, following the interpretation of responds differently to NO_2, which contradicts the expectation that R^*_{GF} as a should serve as a consistent proxy for VOC origin, we would expect more similarity. The major difference . The key difference in our dataset is that CHOCHO dSCDs increase more rapidly with NO_2 in Athens compared to Incheon. The discrepancy hints at different ratios between the emissions of than in Incheon.

In general, increasing NOx and VOCs at these locations. This can be supported by the relative contributions of the top 5 sectors concentrations enhance the formation of HCHO and CHOCHO by promoting the recycling of HOx radicals ($HO_2 + NO \longrightarrow OH + NO_2$), which increases OH concentrations and thus strengthens the oxidation of VOCs. This leads to higher production of both species (Seinfeld and Pandis, 2006). The intrinsic yield of CHOCHO and HCHO from VOC oxidation pathways is generally independent of NOx meaning that R^*_{GF} should remain constant assuming that all ambient factors are the same (e.g., VOC composition, temperature, solar radiation, vertical mixing). However, in high-NOx environments, NOx can suppress OH (through the reaction $OH + NO_2 \longrightarrow HNO_3$), reducing the overall oxidation capacity. This may limit the production of both HCHO and CHOCHO, but depending on their reactivity, this could lead to an increase in R^*_{GF} even if production is reduced.

The discrepancy between Athens and Incheon hints at a different VOC mixture at each location. In Athens, CHOCHO dSCDs increase more rapidly with NO_2 than HCHO dSCDs. A different VOC composition translates via OH-initiated oxidation, followed by reaction with NO, to formation of different alkoxy radicals (RO). The fragmentation of these radicals depends on their structure: Larger alkoxy radicals (e.g., from VOCs like aromatics or alkenes) can fragment into both CHOCHO and HCHO. Contrary, smaller alkoxy radicals (e.g., from methoxy, CH_3O) produce only HCHO. It should be noted that glyoxal is primarily formed from VOCs with double bonds (C=C), such as aromatics, alkenes, and isoprene. Therefore, higher CHOCHO production relative to HCHO (and thus higher R^*_{GF}) suggests a greater contribution from VOCs that produce glyoxal, such as aromatics or unsaturated hydrocarbons, rather than simple alkanes.

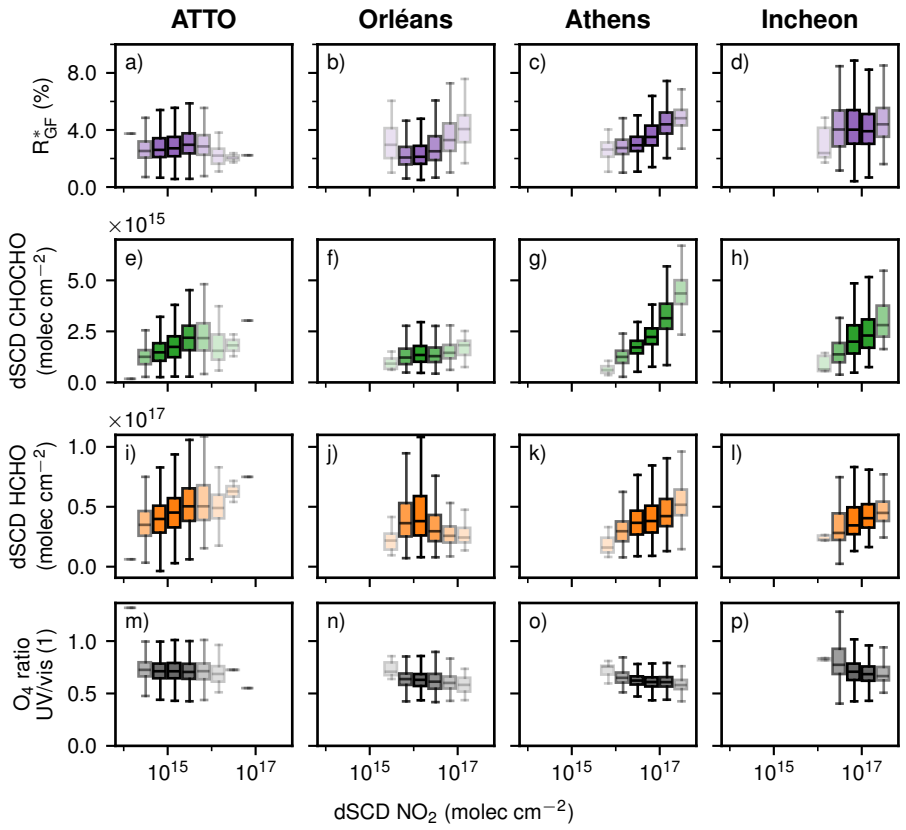

Figure 11. R^*_{GF} (top row), CHOCHO dSCD (upper centre row), HCHO dSCD (lower centre row), and O_4 ratio (bottom row) as a function of binned NO_2 dSCDs for ATTO, Orléans, Athens, and Incheon. Within each box, the horizontal line indicates the median and the box spans the IQR; whiskers extend to 1.5 IQR. Box transparency scales with the number of contributing measurements, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval. Panels e) and i) are self-created based on Donner (2024).

Checking the top five sectors contributing to the total ~~NMVOC emissions from CAMS-GLOB-ANT~~ non-methane VOCs emissions in both cities ~~, see~~ from CAMS-GLOB-ANT (Fig. 12~~. Emissions are similar from the solvents sector and the road transport . The biggest difference between both sites is the large contribution by industrial processes in Incheon~~) shows similar contributions from solvents and ~~the strong contributions~~ road transport but differences in other sectors. Industrial processes dominate in Incheon, whereas refineries and fugitive emissions are more prominent in Athens. Assuming a consistent VOC composition per sector, regardless of the location, the higher CHOCHO emissions in Athens imply that emissions from refineries and ~~fugitives in Athens.~~ fugitive emissions would produce more CHOCHO relative to HCHO than industrial emissions. Possible species with high CHOCHO yields include aromatics (Chan Miller et al., 2016) or acetylene/ethylene (Fu et al., 2008). However, for aromatics, Nishino et al. (2010) found that CHOCHO yield decreases with increasing NO_2

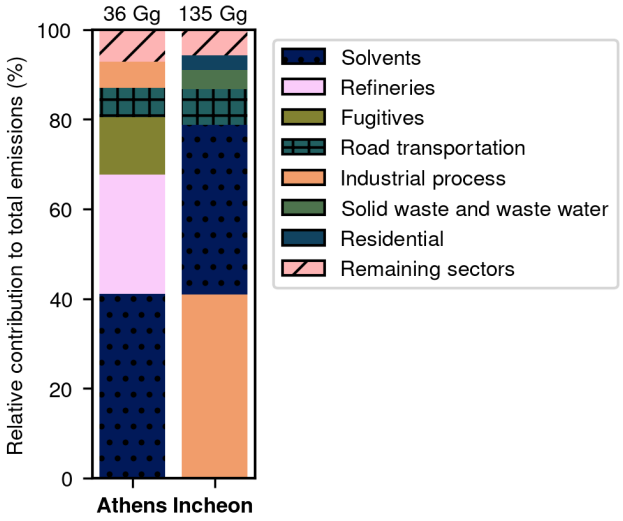

Figure 12. Relative contributions to the CAMS ~~NMVOC~~ non-methane VOCs emissions of the top 5 sectors in Athens and Incheon. All remaining sectors are summarized in one element. ~~Colormap~~ The color map is taken from Crameri et al. (2020).

levels. There is also the possibility, that the declining NO_2 levels in Incheon (Seo et al., 2021) during the measurement period lead to a more stable R_{GF}–NO_2 relationship. But as the Incheon dataset only covers one year, the effect should be minimal.

To summarize, R^*_{GF} shows an inconsistent behaviour with changing NO_2 levels and differs between anthropogenic sites. This implies that (1) systematic R^*_{GF} differences cannot be reduced only to differing NO_2 levels; R^*_{GF} carries additional environmental information. And (2) local factors strongly influence R^*_{GF}, so using it as a proxy for VOC sources likely requires site-specific considerations.

3.4 Comparability between different R_{GF}

~~Measurements of R_{GF} reported in the literature lead to varying and partially contradicting results. This section examines effects that may hinder comparisons between R_{GF} from different measurement techniques. We focus on four key effects and discuss how each is expected to influence the R_{GF} values.~~

3.4.1 Measurement volume

R_{GF} has been computed from data gathered by various different platforms and with different measurement techniques since its first usage. Table 2 ~~shows an overview of the different computations found in literature. Volume~~ lists the various approaches to compute R_{GF}: volume mixing ratios (VMRs), dSCDs with correction terms, and mean VCDs~~have been used to compute R_{GF}. All of~~. All these quantities represent R_{GF} in a different measurement volume. For the particular case of VMR R_{GF} and satellite column-averaged R_{GF}, DiGangi et al. (2012) discusses possible causes for disagreements and also briefly mentions

Table 2. List of different ways to compute R_{GF}

Reference	Method	R_{GF} computation
Vrekoussis et al. (2010)	GOME-2	Annual mean VCDs
DiGangi et al. (2012)	In-situ	surface VMRs[1]
Kluge et al. (2020)	Aircraft	NEMRs[2]
Lerot et al. (2023)	TROPOMI	dSCDs multiplied with NO_2 ratio
Hong et al. (2024)	MAX-DOAS	VMRs removing primary HCHO
This study	MAX-DOAS	dSCDs multiplied with O_4 ratio

[1] volume mixing ratio

[2] normalised excess mixing ratio

the topic of different measurement volumes. We want to further generalize and expand on this inherent difference between the measurement techniques.

Firstly, VMRs obtained by in-situ measurements determine the R_{GF} at the position of the instrument at the sampling time. Here R_{GF} represents the smallest measurement volume, a point measurement.

For R_{GF} values computed via dSCDs from a low elevation angle with O_4 correction (this work), the situation is similar to R_{GF} via VMRs. However, a different volume is probed. Looking towards the horizon, the retrieved dSCDs are dominated by absorption in the lowest layer. Therefore, the resulting R_{GF} is dominated by the volume along the average light path close to the surface until the scattering point.

Lastly, there is column-averaged R_{GF} from either ground-based instruments or satellite-based instruments. Both platforms allow to probe the whole atmospheric column, however with different vertical sensitivities, see Sect. 3.4.2. The column-averaged R_{GF} represents the whole column including the vertical information about the trace gases. However, satellite columns are obtained for the whole ground pixel area, which is larger than the inherent spatial averaging for ground-based columns due to the field of view (FOV).

So even though, all ratios of CHOCHO and HCHO are called R_{GF}, they do not necessarily represent the same measurement volume. Different measurement volumes go along with different kinds of averaging or no averaging at all in the case of in-situ R_{GF}. ~~For the comparability of R_{GF}, it should be noted that~~ Therefore, processes of different scales (spatial or temporal) contribute differently to the R_{GF} ~~from~~ representing different measurement volumes.

3.4.2 Vertical sensitivity

As discussed in the validation study of the TROPOMI HCHO product using ground-based MAX-DOAS observations by De Smedt et al. (2021), satellites and ground-based MAX-DOAS instruments have opposite vertical sensitivity profiles. Satellite-based instruments have minimal sensitivity near the surface, whereas MAX-DOAS instruments are most sensitive at the surface, with sensitivity decreasing to near zero above approximately 3 km altitude. De Smedt et al. (2021) found that accounting for these sensitivity differences can reduce the bias between the two platforms by up to 20%.

For R_{GF}, this implies that satellite-derived values are biased toward higher atmospheric layers compared to ground-based measurements, even when vertical profiles or vertical column densities (VCDs) are used. Notably, previous studies have shown that R_{GF} can vary with altitude. For example, Xing et al. (2020) demonstrated that the diurnal behaviour of R_{GF} changes significantly within the lowest 1 km, which may help explain some discrepancies between satellite and ground-based observations.

3.4.3 ~~Time dependence~~Temporal sampling

Pronounced diurnal and seasonal cycles in R^{*}_{GF} are visible at anthropogenic sites in our dataset. In the presence of such cycles, the time/period of measurement becomes critical. For short duration campaigns, the seasonal cycle has to be considered to avoid a sampling bias.

The diurnal cycle plays an important role when intercomparing satellites or comparing satellites with ground-based instruments. Sun-synchronous low-Earth orbit satellites, such as the one hosting the TROPOMI instrument, pass at a fixed local solar time over the equator and thus only capture a snapshot of the diurnal variability. Given that our observed diurnal cycles are relatively symmetric around noon and the overpass times surround noon (see Fig. ~~??~~5), only minor differences are expected between commonly used satellite instruments such as GOME-2, SCIAMACHY, TROPOMI, and OMI. Only for Athens, the diurnal cycle is shifted to earlier hours, so a notable effect is observed: the measurements during morning overpass are ~~biased~~ higher by approximately 0.5 $\%\mathrm{pt}$. than the afternoon.

When comparing satellite measurements to ground-based instruments, however, systematic differences can emerge in daily averages. In the most extreme case, for Incheon, this could result in an overestimation by TROPOMI of about 0.5 $\%\mathrm{pt}$. relative to the daily average. Importantly, since diurnal variability is most pronounced at anthropogenic sites, the magnitude of this effect differs across environments. Consequently, a spatially variable bias is expected between studies relying solely on satellite data and those based on ground-based observations. When directly comparing both platforms, it is important to use only data close to the overpass time to eliminate this bias. It is worth noting that new and upcoming geostationary satellites (e.g., GEMS, TEMPO, Sentinel-4) provide diurnal coverage, which should help eliminate such biases when comparing R_{GF} from different platforms.

3.4.4 Impact of averaging-ratioing order

~~Next to R_{GF} being computed for different measurement volumes, there are~~ In the literature, one can find two different methodologies ~~found in the literature for the computation of~~ to computation R_{GF} values. Firstly, R_{GF} as the mean of the individual ratios (in the following called instantaneous R_{GF}) and, secondly, R_{GF} as the ratio of the mean of the HCHO and CHOCHO columns (in the following called global R_{GF}).

Both approaches can be applied to any aggregated dataset, but in practise the global ratio is often used for R_{GF} based on satellite data. Satellite retrievals are more challenging than ground-based retrievals: the increased distance to surface-level trace gases and the satellite viewing geometry result in lower sensitivity near the surface (Lerot et al., 2021) and the short integration time limits the signal to noise ratio of the individual measurement. To improve the signal-to-noise ratio, satellite measurements

are commonly averaged over a defined period and area (Lerot et al., 2021) before calculating R_{GF} from the averaged VCDs. The instantaneous R_{GF} is primarily applied for datasets from ground-based instruments, as in this work, since the ground-based instruments generally provide a higher signal-to-noise ratio due to a longer integration time.

The order of operations matters as the division and the mean do not commute in general ~~. So~~ ($\sum \frac{A}{B} \neq \frac{\sum A}{\sum B}$), so the ratio of means (global R_{GF}, Eq. (9)) is not the mean of ratios ~~. Equation (??) shows this for two datasets $A = \{1,2\}$ and $B = \{3,4\}$ in a simplified form, i.e. taking the sum instead of the mean.~~

~~$$\sum \frac{A}{B} = \frac{1}{3} + \frac{2}{4} = \frac{5}{6} \neq \frac{3}{7} = \frac{1+2}{3+4} = \frac{\sum A}{\sum B}$$~~

~~Coming back to~~ (instantaneous R_{GF}~~: Both ratios are shown in~~ , Eq. (8)~~and Eq. (9)~~). Here N refers to the number of CHOCHO dSCDs and M refers to the number of HCHO dSCDs~~; N and M are not necessarily identical.~~

~~$$R_{\mathrm{GF}}^{\mathrm{instantaneous}} = \frac{1}{N} \sum_{i}^{N} \frac{\mathrm{dSCD}_i^{\mathrm{CHOCHO}}}{\mathrm{dSCD}_i^{\mathrm{HCHO}}}$$~~

~~$$R_{\mathrm{GF}}^{\mathrm{global}} = \frac{\frac{1}{N}\sum_i^N \mathrm{dSCD}_i^{\mathrm{CHOCHO}}}{\frac{1}{M}\sum_i^M \mathrm{dSCD}_i^{\mathrm{HCHO}}}$$~~

~~Another interesting aspect is the application of quality filters.~~ The instantaneous R_{GF} requires pairs of simultaneous CHOCHO and HCHO measurements ($M = N$), therefore ensuring direct comparability but reducing data coverage. The global R_{GF} is more forgiving and allows filtering every trace gas individually ($M \neq N$), leading to potential sampling biases. If valid CHOCHO data occur mainly in summer while valid HCHO measurements are available throughout the whole year, the resulting global R_{GF} would mix a seasonal average with an annual average and thus misrepresent the true annual relationship of CHOCHO and HCHO.

$$R_{\mathrm{GF}}^{\mathrm{instant}} = \frac{1}{N} \sum_{i}^{N} \frac{\mathrm{dSCD}_i^{\mathrm{CHOCHO}}}{\mathrm{dSCD}_i^{\mathrm{HCHO}}} \tag{8}$$

$$R_{\mathrm{GF}}^{\mathrm{global}} = \frac{\frac{1}{N}\sum_i^N \mathrm{dSCD}_i^{\mathrm{CHOCHO}}}{\frac{1}{M}\sum_i^M \mathrm{dSCD}_i^{\mathrm{HCHO}}} \tag{9}$$

As the usage of the global R_{GF} is required for practical reasons, we investigate how both approaches differ by applying both methodologies to our ground-based dataset. The quality filters are applied in a way, consistent with the previous ~~chapters~~sections, that only valid pairs of simultaneous CHOCHO and HCHO measurements ($M = N$) are considered for the analysis.

The instantaneous R_{GF} consistently yields higher values compared to the global R_{GF} across all analyses ~~in this study, but the largest discrepancies occur in hourly binned data, see Fig. 13. It shows the diurnal cycle of R_{GF} (without O4 correction) computed with both methodologies.~~ made during this study. A clear systematic bias is visible for the ~~two methods~~differences between the instantaneous R_{GF} and the global R_{GF} in Fig. 13, and the magnitude of this bias varies depending on the station, month, and time of day.

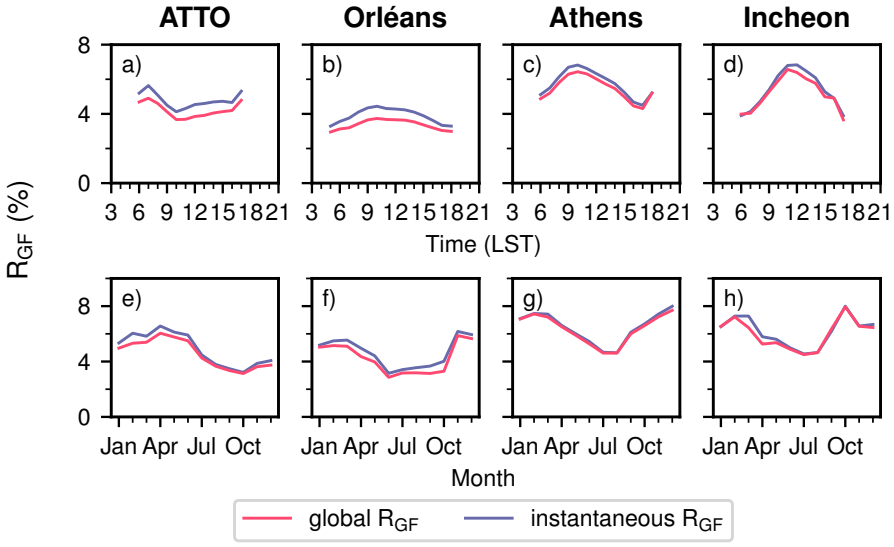

Figure 13. Diurnal cycles (top row) and seasonal cycles (bottom row) of R_{GF} (without O_4 correction) for all four sites. ~~Diurnal cycles of R_{GF} (without correction) for ATTO a), Orléans b), Incheon c) and Athens d) relative to local solar time (LST). The purple line with round markers corresponds to instantaneous R_{GF} values and the red line with the triangle markers corresponds to global R_{GF} values. The mean bias over the day is shown in the legend. Only data from the lowest viewing elevation for every station is included.~~

~~At the ATTO Tower~~ Looking at the diurnal cycles, a large systematic difference is present at ATTO (Fig. 13a) ~~, a large systematic bias is present throughout the entire~~ throughout the day. The ~~strongest difference is observed~~ largest differences occur in Orléans (Fig. 13b), ~~with the largest discrepancies~~ where discrepancies reach just below 1 %pt. ~~occurring around 10am~~around 10:00 local solar time. In contrast, ~~at the anthropogenic stations~~ the anthropogenic sites Incheon and Athens (Fig. 13c+d) ~~,~~ show much closer agreement between the two approaches~~yield more similar results, with smaller differences overall.~~ , with overall smaller differences. A more detailed view of the 10:00 bin distributions, as well as extended daily time series for Orléans, is provided in the supplement (Figs. A9 and A10).

~~Diving into the 10am bin in Orléans and examining the underlying distributionsallows highlighting the differences. As illustrated in Fig. ??, the instantaneous R_{GF} directly incorporates the full distributions of and dSCDs. By calculating the ratio for each individual measurement, a resulting distribution of R_{GF} values (shown in light purple on the right)is obtained, from which the instantaneous R_{GF} value is computed. In contrast, to compute the global R_{GF}, the~~ The magnitude of the difference between both methods depends primarily on the variability and shape of the underlying HCHO and CHOCHO ~~dSCD distributions are first averaged and then the averaged dSCDs are used to compute the mean R_{GF} value.~~

~~Because the underlying distributions and their shapes depend on the data points~~ distributions within each bin, ~~the bias between the two approaches varies across different bins. However, in all our tests, the ground processing consistently yielded higher~~ as well as on their correlation. The consistently higher instantaneous R_{GF} values ~~. This can be explained by the occurrence of~~ are driven by small HCHO dSCDs in the denominator, which ~~amplify the calculated ratio. To put simply,~~

~~the difference depends on the~~ disproportionately increase individual ratios and introduce skewness. Consequently, the bias between both methods increases with growing asymmetry of the ~~R_{GF} distribution, as greater skewness shifts the instantaneous R_{GF} further away from the global R_{GF} mean. However, this simplified view ignores the interactions, as the skewness of the R_{GF} distribution depends on the underlying and dSCD distributions and ~~correlation in the bin.~~ ratio distribution.~~ For a more ~~technical mathematical argument under which circumstances~~ formal reasoning of the conditions under which the ratio of means ~~can be identical to~~ equals the mean of ratios ~~for two random variables,~~ see Heijmans (1999).

~~Visualization of the different averaging approaches using the 10am bin from the lowest viewing elevation from Orléans. The distributions of dSCD (top) and of dSCD (bottom) are shown on the left, with the respective mean indicated with the red vertical line. On the right, the distribution of R_{GF} and the instantaneous R_{GF} (purple vertical line) is shown. The vertical red line corresponds to the global R_{GF}.~~

4 Summary and c~~C~~onclusions

Over the past decade, the literature has reported multiple inconclusive findings regarding the ratio of glyoxal-to-formaldehyde, R_{GF}, and its use as a proxy for VOC source identification. In this study, we use a multi-year ground-based MAX-DOAS dataset at four stations to revisit R_{GF} and reassess its drivers and limitations. Our dataset includes four MAX-DOAS stations located in different environments, allowing us to systematically investigate patterns in the data. Additionally, we compare the results with various meteorological variables and other trace gases.

We find differences in the absolute magnitudes of R^*_{GF} across environments: lower values at the biogenic sites (ATTO and Orléans) and high values at the strongly anthropogenic sites (Incheon and Athens). While the dSCDs of CHOCHO and HCHO are similarly high across all stations, both trace gases show different behaviours. Glyoxal is notably enhanced at the anthropogenic sites and serves as the primary factor driving the differences in absolute R^*_{GF} magnitudes. This offset is consistently observed in the seasonal, weekly, and diurnal cycles.

In addition, we observe a seasonal cycle characterised by higher R^*_{GF} values during the Northern Hemisphere winter months and lower values during the summer months, primarily driven by the pronounced seasonal cycle of HCHO, which is strongly linked to temperature. This pattern holds across all stations except ATTO, where the seasonal cycle is shifted by several months and exhibits enhanced values during the wet and reduced values during the dry season (Donner, 2024).

The diurnal cycles of R^*_{GF} are relatively flat at the biogenic stations but pronounced at the anthropogenic stations, showing variations of approximately 2 $\%\mathrm{pt}.$ with peaks occurring before or around noon. Glyoxal, in particular, exhibits a distinct diurnal pattern at the anthropogenic sites compared to the biogenic sites, making it the main driver of the observed diurnal R^*_{GF} variability.

We also detect a weak weekend effect at the anthropogenic stations, with R^*_{GF} decreasing by about 10 $\%$ compared to weekdays, primarily due to a stronger reduction by CHOCHO on the weekend.

Moreover, we ~~report~~ investigated the link to meteorology of R^*_{GF}. It exhibits a clear temperature dependence~~of R^*_{GF}~~. Above approximately 15 °C, HCHO levels increase more strongly with temperature than CHOCHO, leading to a hyperbolic decrease

in R^*_{GF} beyond this temperature. Accounting for the temperature dependence with a regression-based approach removes all variation of R^*_{GF} with other meteorological variables due to their intercorrelation. The seasonal cycle of R^*_{GF} for non-tropical sites is removed, whereas the diurnal cycle remains.

Our investigation of the dependence of ~~R^*_{GF} on~~ the R_{GF}–$\mathrm{NO_2}$ relationship implies that R^*_{GF} cannot be ~~reduced~~ linked only to different $\mathrm{NO_2}$ levels. Local effects and VOC emission characteristics may play a more significant role than previously assumed. Thus site-specific considerations for the usage of R^*_{GF} as a proxy for VOC type might be required. Overall, both the diurnal cycles and the relationship to $\mathrm{NO_2}$ suggest that CHOCHO ~~seems to be stronger connected to~~ is more closely associated with anthropogenic activity than HCHO.

Finally, we examine four factors that can impact comparisons of R_{GF} from different datasets. Firstly, R_{GF}, computed from different measurement techniques, inherently average over different measurement volumes. In addition, the vertical sensitivity of satellites and ground-based instruments is not identical. Moreover, due to the pronounced seasonal and diurnal cycles, the time of the measurement becomes critical. No significant difference between different overpass times of GOME-2, SCIAMACHY, and TROPOMI/OMI is observed due to the symmetric diurnal cycles~~,~~; however the overpass time has to be considered for ground-based and satellite comparisons. Lastly, we investigate the impact of the order of ratioing and averaging. The global R_{GF} is generally biased low compared to the instantaneous R_{GF} as no extreme values occur in the denominator.

~~Further work is needed to expand our understanding and~~ Future work should quantify these effects through radiative transfer simulations or direct instrument intercomparisons. Beyond this, expanding the network of long-term ground-based observations would strengthen the statistical basis ~~through additional multi-year ground-based observations~~and broaden the range of environments and source regimes captured, ultimately advancing the interpretation of R_{GF} as a proxy for VOC origin. Improved coverage of simultaneous VOC precursor measurements would further enhance the interpretation of observational data. An important complementary perspective could come from modeling studies. In particular, sensitivity studies using chemical box models (including gas-phase and heterogeneous chemistry) offer a unique opportunity to challenge current understanding. By systematically turning individual processes on or off, such models can help disentangle complex interconnections that are difficult to isolate in observational data. Additionally, distinguishing between the primary and secondary formation of HCHO and CHOCHO in the context of R_{GF} could help clarify the underlying mechanisms (Hong et al., 2024; Xing et al., 2025).

Given these complexities, using the glyoxal-to-formaldehyde ratio as a proxy for VOC source identification remains challenging. While we observe clear differences in the absolute magnitudes of R^*_{GF} across stations in different environments, suggesting that the ratio carries valuable environmental information, our incomplete understanding of emissions, VOC precursors, and the detailed chemistry of HCHO and CHOCHO currently limits the reliable use of R^*_{GF} as a robust proxy for VOC source attribution.

Data availability. The ERA5 atmospheric reanalysis data were downloaded from the from the Copernicus Climate Change (C3S) climate data store (CDS): https://doi.org/10.24381/cds.adbb2d47 (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023). The

emission datasets, CAMS-GLOB-ANT version 6.2 and CAMS-GLOB-BIO version 3.1, were downloaded from ECCAD: https://permalink.aeris-data.fr/CAMS-GLOB-ANT (Granier et al., 2019). MAX-DOAS data are available from the authors on request.

Author contributions. SB, AR, and MV contributed to the conceptualization, methodology, data interpretation and writing of the original draft. SB and BZ set-up and operated the instrument in Orléans. SD and TW conducted the measurements and provided the ATTO data set. APP performed the backwards sensitivity studies. LA provided expertise regarding the fit settings for the Bremen instruments. SB processed the data and performed the data analysis. All authors have contributed with scientific discussions to data interpretation.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Atmospheric Chemistry and Physics.

Disclaimer. The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

Acknowledgements. AI Usage: Reformatting tables, ~~Improving~~ improving code for figures, indexing literature, ~~Improved~~ improved text readability, and spelling correction.

Simulations were performed on the HPC cluster Aether at the University of Bremen, financed by the Deutsche Forschungsgemeinschaft (DFG) within the scope of the Excellence Initiative

~~The~~. The authors thank Morgan Lopez, Dylan Lopez, and Christof Petri for their support in maintaining the Traînou (Orléans) measurement site. The ATTO site is operated and managed by INPA and MPG. The authors also acknowledge the whole ATTO team for maintenance and operation of the ATTO site and its infrastructure. Further, Bianca Lauster and Steffen Ziegler are acknowledged for their large contribution in operating and maintaining the ATTO MAX-DOAS instrument. The authors thank Myrto Gratsea and NOA for supporting the station in Athens and the GMAP and SIJAQ teams for facilitating measurements in Incheon. The ATTO research has been supported by the German Federal Ministry of Education and Research (BMBF contract 01LK2101B).

Measurements in Incheon have received funding from NIER. SB, MV, AR acknowledge financial support from the University of Bremen. SB acknowledge financial support from the CLIMADEMY ERASMUS+ (101056066) Grant. MV ~~acknowledges~~ and APP acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (University Allowance, EXC 2077, University of Bremen). This work was supported by the Data Science Center of the University of Bremen (DSC@UB) funded by the State of Bremen.

References

Alvarado, L. M. A., Richter, A., Vrekoussis, M., Hilboll, A., Hedegaard, A. B. K., Schneising, O., and Burrows, J. P.: Unexpected long-range transport of glyoxal and formaldehyde observed from the Copernicus Sentinel-5 Precursor satellite during the 2018 Canadian wildfires, Atmospheric Chemistry and Physics, 20, 2057–2072, https://doi.org/10.5194/acp-20-2057-2020, 2020.

Andreae, M. O. and Merlet, P.: Emission of trace gases and aerosols from biomass burning, Global Biogeochemical Cycles, 15, 955–966, https://doi.org/10.1029/2000gb001382, 2001.

Andreae, M. O., Acevedo, O. C., Araùjo, A., Artaxo, P., Barbosa, C. G. G., Barbosa, H. M. J., Brito, J., Carbone, S., Chi, X., Cintra, B. B. L., da Silva, N. F., Dias, N. L., Dias-Júnior, C. Q., Ditas, F., Ditz, R., Godoi, A. F. L., Godoi, R. H. M., Heimann, M., Hoffmann, T., Kesselmeier, J., Könemann, T., Krüger, M. L., Lavric, J. V., Manzi, A. O., Lopes, A. P., Martins, D. L., Mikhailov, E. F., Moran-Zuloaga, D., Nelson, B. W., Nölscher, A. C., Santos Nogueira, D., Piedade, M. T. F., Pöhlker, C., Pöschl, U., Quesada, C. A., Rizzo, L. V., Ro, C.-U., Ruckteschler, N., Sá, L. D. A., de Oliveira Sá, M., Sales, C. B., dos Santos, R. M. N., Saturno, J., Schöngart, J., Sörgel, M., de Souza, C. M., de Souza, R. A. F., Su, H., Targhetta, N., Tóta, J., Trebs, I., Trumbore, S., van Eijck, A., Walter, D., Wang, Z., Weber, B., Williams, J., Winderlich, J., Wittmann, F., Wolff, S., and Yáñez Serrano, A. M.: The Amazon Tall Tower Observatory (ATTO): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols, Atmospheric Chemistry and Physics, 15, 10 723–10 776, https://doi.org/10.5194/acp-15-10723-2015, 2015.

Beirle, S., Platt, U., Wenig, M., and Wagner, T.: Weekly cycle of NO_2 by GOME measurements: a signature of anthropogenic sources, Atmospheric Chemistry and Physics, 3, 2225–2232, https://doi.org/10.5194/acp-3-2225-2003, 2003.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., and Dominici, F.: Ozone and Short-term Mortality in 95 US Urban Communities, 1987-2000, JAMA, 292, 2372–2378, https://doi.org/10.1001/jama.292.19.2372, 2004.

Bello, J., Vallet, P., Perot, T., Balandier, P., Seigner, V., Perret, S., Couteau, C., and Korboulewsky, N.: How do mixing tree species and stand density affect seasonal radial growth during drought events?, Forest Ecology and Management, 432, 436–445, https://doi.org/10.1016/j.foreco.2018.09.044, 2019.

Berg, F., Novelli, A., Dubus, R., Hofzumahaus, A., Holland, F., Wahner, A., and Fuchs, H.: Temperature-dependent rate coefficients for the reactions of OH radicals with selected alkanes, aromatic compounds, and monoterpenes, Atmospheric Chemistry and Physics, 24, 13 715–13 731, https://doi.org/10.5194/acp-24-13715-2024, 2024.

Bourtsoukidis, E., Pozzer, A., Williams, J., Makowski, D., Peñuelas, J., Matthaios, V. N., Lazoglou, G., Yañez-Serrano, A. M., Lelieveld, J., Ciais, P., Vrekoussis, M., Daskalakis, N., and Sciare, J.: High temperature sensitivity of monoterpene emissions from global vegetation, Communications Earth & Environment, 5, https://doi.org/10.1038/s43247-023-01175-9, 2024.

Castro-Melgar, I., Tsagkou, A., Zacharopoulou, M., Basiou, E., Athinelis, I., Katris, E.-A., Kalavrezou, I.-E., and Parcharidis, I.: Wildfires During Early Summer in Greece (2024): Burn Severity and Land Use Dynamics Through Sentinel-2 Data, Forests, 16, 268, https://doi.org/10.3390/f16020268, 2025.

Chan Miller, C., Gonzalez Abad, G., Wang, H., Liu, X., Kurosu, T., Jacob, D. J., and Chance, K.: Glyoxal retrieval from the Ozone Monitoring Instrument, Atmospheric Measurement Techniques, 7, 3891–3907, https://doi.org/10.5194/amt-7-3891-2014, 2014.

Chan Miller, C., Jacob, D. J., González Abad, G., and Chance, K.: Hotspot of glyoxal over the Pearl River delta seen from the OMI satellite instrument: implications for emissions of aromatic hydrocarbons, Atmospheric Chemistry and Physics, 16, 4631–4639, https://doi.org/10.5194/acp-16-4631-2016, 2016.

Chan Miller, C., Jacob, D. J., Marais, E. A., Yu, K., Travis, K. R., Kim, P. S., Fisher, J. A., Zhu, L., Wolfe, G. M., Hanisco, T. F., Keutsch, F. N., Kaiser, J., Min, K.-E., Brown, S. S., Washenfelder, R. A., Abad, G. G., and Chance, K.: Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data, Atmospheric Chemistry and Physics, 17, 8725–8738, https://doi.org/10.5194/acp-17-8725-2017, 2017.

Chance, K. and Kurucz, R.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 1289–1295, https://doi.org/10.1016/j.jqsrt.2010.01.036, 2010.

Chen, Y., Liu, C., Su, W., Hu, Q., Zhang, C., Liu, H., and Yin, H.: Identification of volatile organic compound emissions from anthropogenic and biogenic sources based on satellite observation of formaldehyde and glyoxal, Science of The Total Environment, 859, 159 997, https://doi.org/10.1016/j.scitotenv.2022.159997, 2023.

Copernicus Climate Change Service (C3S) Climate Data Store (CDS): ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/cds.adbb2d47, last access: 10 March 2026, 2023.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nature Communications, 11, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Danckaert, T., Fayt, C., van Roozendael, M., de Smedt, I., Letocart, V., Merlaud, A., and Pinardi, G.: QDOAS Software user manual - Version 3.2, Royal Belgian Institute for Space Aeronomy, https://uv-vis.aeronomie.be/software/QDOAS/QDOAS_manual.pdf, last access: 10 March 2026, 2017.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C., Romahn, F., Ryan, R. G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., and Van Roozendael, M.: Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements, Atmospheric Chemistry and Physics, 21, 12 561–12 593, https://doi.org/10.5194/acp-21-12561-2021, 2021.

Delacre, M., Lakens, D., and Leys, C.: Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test, International Review of Social Psychology, 30, 92–101, https://doi.org/10.5334/irsp.82, 2017.

Delacre, M., Leys, C., Mora, Y. L., and Lakens, D.: Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's F-test instead of the Classical F-test in One-Way ANOVA, International Review of Social Psychology, 32, 13, https://doi.org/10.5334/irsp.198, 2019.

Derwent, R. G., Jenkin, M. E., Utembe, S. R., Shallcross, D. E., Murrells, T. P., and Passant, N. R.: Secondary organic aerosol formation from a large number of reactive man-made organic compounds, Science of The Total Environment, 408, 3374–3381, https://doi.org/10.1016/j.scitotenv.2010.04.013, 2010.

Dienhart, D., Crowley, J. N., Bourtsoukidis, E., Edtbauer, A., Eger, P. G., Ernle, L., Harder, H., Hottmann, B., Martinez, M., Parchatka, U., Paris, J.-D., Pfannerstill, E. Y., Rohloff, R., Schuladen, J., Stönner, C., Tadic, I., Tauer, S., Wang, N., Williams, J., Lelieveld, J., and Fischer, H.: Measurement report: Observation-based formaldehyde production rates and their relation to OH reactivity around the Arabian Peninsula, Atmospheric Chemistry and Physics, 21, 17 373–17 388, https://doi.org/10.5194/acp-21-17373-2021, 2021.

DiGangi, J. P., Henry, S. B., Kammrath, A., Boyle, E. S., Kaser, L., Schnitzhofer, R., Graus, M., Turnipseed, A., Park, J.-H., Weber, R. J., Hornbrook, R. S., Cantrell, C. A., Maudlin III, R. L., Kim, S., Nakashima, Y., Wolfe, G. M., Kajii, Y., Apel, E. C., Goldstein, A. H., Guenther, A., Karl, T., Hansel, A., and Keutsch, F. N.: Observations of glyoxal and formaldehyde as metrics for the anthropogenic impact on rural photochemistry, Atmospheric Chemistry and Physics, 12, 9529–9543, https://doi.org/10.5194/acp-12-9529-2012, 2012.

Donner, S.: Retrieving vertical profiles and tropospheric columns of formaldehyde from global long-term MAX-DOAS measurements, Ph.D. thesis, Johannes Gutenberg-Universität Mainz, https://doi.org/10.25358/OPENSCIENCE-11085, 2024.

Ehhalt, D., Prather, M., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., Matson, P., Midgley, P., and Wang, M.: Atmospheric Chemistry and Greenhouse Gases. In: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. (Eds.), Cambridge University Press, ISBN 0521807670, 2001.

European Environment Agency: Harm to human health from air pollution in Europe: burden of disease status, Briefing no. 21/2024, https://doi.org/10.2800/3950756, 2024.

Finkenzeller, H. and Volkamer, R.: O_2–O_2 CIA in the gas phase: Cross-section of weak bands, and continuum absorption between 297–500 nm, Journal of Quantitative Spectroscopy and Radiative Transfer, 279, 108 063, https://doi.org/https://doi.org/10.1016/j.jqsrt.2021.108063, 2022.

Fleischmann, O. C., Hartmann, M., Burrows, J. P., and Orphal, J.: New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, Journal of Photochemistry and Photobiology A: Chemistry, 168, 117–132, https://doi.org/10.1016/j.jphotochem.2004.03.026, 2004.

Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P., Chance, K., and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-species inversion system, Atmospheric Chemistry and Physics, 12, 6699–6721, https://doi.org/10.5194/acp-12-6699-2012, 2012.

Franco, B., Marais, E. A., Bovy, B., Bader, W., Lejeune, B., Roland, G., Servais, C., and Mahieu, E.: Diurnal cycle and multi-decadal trend of formaldehyde in the remote atmosphere near 46N, Atmospheric Chemistry and Physics, 16, 4171–4189, https://doi.org/10.5194/acp-16-4171-2016, 2016.

Fu, T.-M., Jacob, D. J., Wittrock, F., Burrows, J. P., Vrekoussis, M., and Henze, D. K.: Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009505, 2008.

Games, P. A. and Howell, J. F.: Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study, Journal of Educational Statistics, 1, 113–125, https://doi.org/10.3102/10769986001002113, 1976.

Gilman, J. B., Lerner, B. M., Kuster, W. C., Goldan, P. D., Warneke, C., Veres, P. R., Roberts, J. M., de Gouw, J. A., Burling, I. R., and Yokelson, R. J.: Biomass burning emissions and potential air quality impacts of volatile organic compounds and other trace gases from fuels common in the US, Atmospheric Chemistry and Physics, 15, 13 915–13 938, https://doi.org/10.5194/acp-15-13915-2015, 2015.

Goldstein, A. H. and Galbally, I. E.: Known and Unexplored Organic Constituents in the Earth's Atmosphere, Environmental Science & Technology, 41, 1514–1521, https://doi.org/10.1021/es072476p, 2007.

Granier, C., Darras, S., Denier van der Gon, H., Doubalova, J., Elguindi, N., Galle, B., Gauss, M., Guevara, M., Jalkanen, J.-P., Kuenen, J., Liousse, C., Quack, B., Simpson, D., and Sindelarova, K.: The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version), https://doi.org/10.24380/D0BN-KX16, 2019.

Gratsea, M., Vrekoussis, M., Richter, A., Wittrock, F., Schönhardt, A., Burrows, J., Kazadzis, S., Mihalopoulos, N., and Gerasopoulos, E.: Slant column MAX-DOAS measurements of nitrogen dioxide, formaldehyde, glyoxal and oxygen dimer in the urban environment of Athens, Atmospheric Environment, 135, 118–131, https://doi.org/10.1016/j.atmosenv.2016.03.048, 2016.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A global model of natural volatile organic compound emissions, Journal of Geophysical Research: Atmospheres, 100, 8873–8892, https://doi.org/10.1029/94jd02950, 1995.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R.: Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses, Journal of Geophysical Research: Atmospheres, 98, 12 609–12 617, https://doi.org/10.1029/93jd00527, 1993.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geoscientific Model Development, 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Guo, Y., Wang, S., Zhu, J., Zhang, R., Gao, S., Saiz-Lopez, A., and Zhou, B.: Atmospheric formaldehyde, glyoxal and their relations to ozone pollution under low- and high-NOx regimes in summertime Shanghai, China, Atmospheric Research, 258, 105 635, https://doi.org/10.1016/j.atmosres.2021.105635, 2021.

Haagen-Smit, A. J.: Chemistry and Physiology of Los Angeles Smog, Industrial & Engineering Chemistry, 44, 1342–1346, https://doi.org/10.1021/ie50510a045, 1952.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmospheric Chemistry and Physics, 9, 5155–5236, https://doi.org/10.5194/acp-9-5155-2009, 2009.

Heijmans, R.: When does the expectation of a ratio equal the ratio of expectations?, Statistical Papers, 40, 107–115, https://doi.org/10.1007/bf02927114, 1999.

Hellenic Statistical Authority: Attiki census results of population and housing - ELSTAT 2021, https://www.statistics.gr/documents/20181/18409455/population_ATTICA_EN.pdf/f8156495-873d-bb66-bf8e-b87bf6974e14, last access: 10 March 2026, 2024.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/CDS.ADBB2D47, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023.

Hong, Q., Xing, J., Xing, C., Yang, B., Su, W., Chen, Y., Zhang, C., Zhu, Y., and Liu, C.: Investigating vertical distributions and photochemical indications of formaldehyde, glyoxal, and NO_2 from MAX-DOAS observations in four typical cities of China, Science of The Total Environment, 954, 176 447, https://doi.org/10.1016/j.scitotenv.2024.176447, 2024.

Hoque, H. M. S., Irie, H., and Damiani, A.: First MAX-DOAS Observations of Formaldehyde and Glyoxal in Phimai, Thailand, Journal of Geophysical Research: Atmospheres, 123, 9957–9975, https://doi.org/10.1029/2018jd028480, 2018a.

Hoque, H. M. S., Irie, H., Damiani, A., Rawat, P., and Naja, M.: First Simultaneous Observations of Formaldehyde and Glyoxal by MAX-DOAS in the Indo-Gangetic Plain Region, SOLA, 14, 159–164, https://doi.org/10.2151/sola.2018-028, 2018b.

Hönninger, G., von Friedeburg, C., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS), Atmospheric Chemistry and Physics, 4, 231–254, https://doi.org/10.5194/acp-4-231-2004, 2004.

Institut national de la statistique et des études économiques: Populations de référence 2022 - Commune d'Orléans (45234), https://www.insee.fr/fr/statistiques/8288323?geo=COM-45234, last access: 10 March 2026, 2024.

Institut national de la statistique et des études économiques: Dossier complet - Commune de Traînou (45327), https://www.insee.fr/fr/statistiques/2011101?geo=COM-45327, last access: 10 March 2026, 2025.

Instituto Brasileiro de Geografia e Estatística: Panorama Censo 2022 - Manaus, https://censo2022.ibge.gov.br/panorama/, last access: 10 March 2026, 2022.

Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, Atmospheric Measurement Techniques, 4, 1027–1044, https://doi.org/10.5194/amt-4-1027-2011, 2011.

Kaiser, J., Wolfe, G. M., Min, K. E., Brown, S. S., Miller, C. C., Jacob, D. J., deGouw, J. A., Graus, M., Hanisco, T. F., Holloway, J., Peischl, J., Pollack, I. B., Ryerson, T. B., Warneke, C., Washenfelder, R. A., and Keutsch, F. N.: Reassessing the ratio of glyoxal to formaldehyde as an indicator of hydrocarbon precursor speciation, Atmospheric Chemistry and Physics, 15, 7571–7583, https://doi.org/10.5194/acp-15-7571-2015, 2015.

Kassomenos, P., Kotroni, V., and Kallos, G.: Analysis of climatological and air quality observations from Greater Athens Area, Atmospheric Environment, 29, 3671–3688, https://doi.org/10.1016/1352-2310(94)00358-r, 1995.

Kim, M.-J., Baek, K.-M., Heo, J.-B., Cheong, J.-P., and Baek, S.-O.: Concentrations, health risks, and sources of hazardous air pollutants in Seoul-Incheon, a megacity area in Korea, Air Quality, Atmosphere & Health, 14, 873–893, https://doi.org/10.1007/s11869-021-00986-z, 2021.

Kluge, F., Hüneke, T., Knecht, M., Lichtenstern, M., Rotermund, M., Schlager, H., Schreiner, B., and Pfeilsticker, K.: Profiling of formaldehyde, glyoxal, methylglyoxal, and CO over the Amazon: normalized excess mixing ratios and related emission factors in biomass burning plumes, Atmospheric Chemistry and Physics, 20, 12 363–12 389, https://doi.org/10.5194/acp-20-12363-2020, 2020.

Kurucz, R. L., Furenlid, I., Brault, J., and Testerman, L.: Solar flux atlas from 296 to 1300 nm, National Solar Observatory Atlas No. 1, 1984.

Lagouvardos, K., Kotroni, V., Giannaros, T. M., and Dafis, S.: Meteorological Conditions Conducive to the Rapid Spread of the Deadly Wildfire in Eastern Attica, Greece, Bulletin of the American Meteorological Society, 100, 2137–2145, https://doi.org/10.1175/bams-d-18-0231.1, 2019.

Lange, K., Richter, A., Bösch, T., Zilker, B., Latsch, M., Behrens, L. K., Okafor, C. M., Bösch, H., Burrows, J. P., Merlaud, A., Pinardi, G., Fayt, C., Friedrich, M. M., Dimitropoulou, E., Van Roozendael, M., Ziegler, S., Ripperger-Lukosiunaite, S., Kuhn, L., Lauster, B., Wagner, T., Hong, H., Kim, D., Chang, L.-S., Bae, K., Song, C.-K., Park, J.-U., and Lee, H.: Validation of GEMS tropospheric NO_2 columns and their diurnal variation with ground-based DOAS measurements, Atmospheric Measurement Techniques, 17, 6315–6344, https://doi.org/10.5194/amt-17-6315-2024, 2024.

Laothawornkitkul, J., Taylor, J. E., Paul, N. D., and Hewitt, C. N.: Biogenic volatile organic compounds in the Earth system, New Phytologist, 183, 27–51, https://doi.org/10.1111/j.1469-8137.2009.02859.x, 2009.

Lee, M., Heikes, B. G., and Jacob, D. J.: Enhancements of hydroperoxides and formaldehyde in biomass burning impacted air and their effect on atmospheric oxidant cycles, Journal of Geophysical Research: Atmospheres, 103, 13 201–13 212, https://doi.org/10.1029/98jd00578, 1998.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367–371, https://doi.org/10.1038/nature15371, 2015.

Lerot, C., Hendrick, F., Roozendael, M. V., Alvarado, L. M. A., Richter, A., Smedt, I. D., Theys, N., Vlietinck, J., Yu, H., Gent, J. V., Stavrakou, T., Müller, J.-F., Valks, P., Loyola, D., Irie, H., Kumar, V., Wagner, T., Schreier, S. F., Sinha, V., Wang, T., Wang, P., and

Retscher, C.: Glyoxal tropospheric column retrievals from TROPOMI - multi-satellite intercomparison and ground-based validation, Atmospheric Measurement Techniques, 14, 7775–7807, https://doi.org/10.5194/amt-14-7775-2021, 2021.

Lerot, C., Müller, J.-F., Theys, N., De Smedt, I., Stavrakou, T., and Van Roozendael, M.: Satellite Evidence for Glyoxal Depletion in Elevated Fire Plumes, Geophysical Research Letters, 50, e2022GL102 195, https://doi.org/10.1029/2022GL102195, 2023.

Li, X., Rohrer, F., Brauers, T., Hofzumahaus, A., Lu, K., Shao, M., Zhang, Y. H., and Wahner, A.: Modeling of HCHO and CHOCHO at a semi-rural site in southern China during the PRIDE-PRD2006 campaign, Atmospheric Chemistry and Physics, 14, 12 291–12 305, https://doi.org/10.5194/acp-14-12291-2014, 2014.

Li, X., Zhu, L., De Smedt, I., Sun, W., Chen, Y., Shu, L., Wang, D., Liu, S., Pu, D., Li, J., Zuo, X., Fu, W., Li, Y., Zhang, P., Yan, Z., Fu, T.-M., Shen, H., Wang, C., Ye, J., and Yang, X.: Global Temperature Dependency of Biogenic HCHO Columns Observed From Space: Interpretation of TROPOMI Results Using GEOS-Chem Model, Journal of Geophysical Research: Atmospheres, 129, e2024JD041 784, https://doi.org/10.1029/2024JD041784, 2024.

Liaskoni, M., Huszár, P., Bartík, L., Prieto Perez, A. P., Karlický, J., and Šindelářová, K.: The long-term impact of biogenic volatile organic compound emissions on urban ozone patterns over central Europe: contributions from urban and rural vegetation, Atmospheric Chemistry and Physics, 24, 13 541–13 569, https://doi.org/10.5194/acp-24-13541-2024, 2024.

Liu, M. and Matsui, H.: Secondary Organic Aerosol Formation Regulates Cloud Condensation Nuclei in the Global Remote Troposphere, Geophysical Research Letters, 49, e2022GL100 543, https://doi.org/10.1029/2022GL100543, 2022.

MacDonald, S. M., Oetjen, H., Mahajan, A. S., Whalley, L. K., Edwards, P. M., Heard, D. E., Jones, C. E., and Plane, J. M. C.: DOAS measurements of formaldehyde and glyoxal above a south-east Asian tropical rainforest, Atmospheric Chemistry and Physics, 12, 5949–5962, https://doi.org/10.5194/acp-12-5949-2012, 2012.

Mavroidis, I. and Ilia, M.: Trends of NOx, NO_2 and O_3 concentrations at three different types of air quality monitoring stations in Athens, Greece, Atmospheric Environment, 63, 135–147, https://doi.org/10.1016/j.atmosenv.2012.09.030, 2012.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth System Science Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

Meller, R. and Moortgat, G. K.: Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 K in the wavelength range 225–375 nm, Journal of Geophysical Research: Atmospheres, 105, 7089–7101, https://doi.org/10.1029/1999jd901074, 2000.

Myriokefalitakis, S., Vrekoussis, M., Tsigaridis, K., Wittrock, F., Richter, A., Brühl, C., Volkamer, R., Burrows, J. P., and Kanakidou, M.: The influence of natural and anthropogenic secondary sources on the glyoxal global distribution, Atmospheric Chemistry and Physics, 8, 4965–4981, https://doi.org/10.5194/acp-8-4965-2008, 2008.

Nelson, P. F., Tibbett, A. R., and Day, S. J.: Effects of vehicle type and fuel quality on real world toxic emissions from diesel vehicles, Atmospheric Environment, 42, 5291–5303, https://doi.org/10.1016/j.atmosenv.2008.02.049, 2008.

Nguyen, H. T., Kim, K.-H., and Park, C.: Long-term trend of NO_2 in major urban areas of Korea and possible consequences for health, Atmospheric Environment, 106, 347–357, https://doi.org/10.1016/j.atmosenv.2015.02.003, 2015.

Nishino, N., Arey, J., and Atkinson, R.: Formation Yields of Glyoxal and Methylglyoxal from the Gas-Phase OH Radical-Initiated Reactions of Toluene, Xylenes, and Trimethylbenzenes as a Function of NO_2 Concentration, 114, 10 140–10 147, https://doi.org/10.1021/jp105112h, 2010.

Nussbaumer, C. M., Crowley, J. N., Schuladen, J., Williams, J., Hafermann, S., Reiffs, A., Axinte, R., Harder, H., Ernest, C., Novelli, A., Sala, K., Martinez, M., Mallik, C., Tomsche, L., Plass-Dülmer, C., Bohn, B., Lelieveld, J., and Fischer, H.: Measurement report: Photochemical production and loss rates of formaldehyde and ozone across Europe, Atmospheric Chemistry and Physics, 21, 18 413–18 432, https://doi.org/10.5194/acp-21-18413-2021, 2021.

Paris, E., Carnevale, M., Vincenti, B., Palma, A., Guerriero, E., Borello, D., and Gallucci, F.: Evaluation of VOCs Emitted from Biomass Combustion in a Small CHP Plant: Difference between Dry and Wet Poplar Woodchips, Molecules, 27, 955, https://doi.org/10.3390/molecules27030955, 2022.

Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, A., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., Hendrick, F., Irie, H., Merlaud, A., Navarro Comas, M., Peters, E., Piters, A. J. M., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vrekoussis, M., Wagner, T., Wittrock, F., and Yilmaz, S.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, 6, 167–185, https://doi.org/10.5194/amt-6-167-2013, 2013.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geoscientific Model Development, 12, 4955–4997, https://doi.org/10.5194/gmd-12-4955-2019, 2019.

Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy: Principles and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-21193-8, https://doi.org/10.1007/978-3-540-75776-4_6, 2008.

Poulidis, A. P. and Takemi, T.: A 1998—2013 climatology of Kyushu, Japan: seasonal variations of stability and rainfall, International Journal of Climatology, 37, 1843–1858, https://doi.org/10.1002/joc.4817, 2016.

Pusede, S. E., Gentner, D. R., Wooldridge, P. J., Browne, E. C., Rollins, A. W., Min, K.-E., Russell, A. R., Thomas, J., Zhang, L., Brune, W. H., Henry, S. B., DiGangi, J. P., Keutsch, F. N., Harrold, S. A., Thornton, J. A., Beaver, M. R., St. Clair, J. M., Wennberg, P. O., Sanders, J., Ren, X., VandenBoer, T. C., Markovic, M. Z., Guha, A., Weber, R., Goldstein, A. H., and Cohen, R. C.: On the temperature dependence of organic reactivity, nitrogen oxides, ozone production, and the impact of emission controls in San Joaquin Valley, California, Atmospheric Chemistry and Physics, 14, 3373–3395, https://doi.org/10.5194/acp-14-3373-2014, 2014.

Ramonet, M., Lopez, M., and Delmotte, M.: ICOS ATC NRT CO_2 growing time series from Trainou (50.0 m), https://hdl.handle.net/11676/bP05gr9ZOJVKz7aARFR0lb7f, last access: 10 March 2026, 2025.

Rawat, P., Naja, M., Rajwar, M. C., Irie, H., Lerot, C., Kumar, M., and Lal, S.: Long-term observations of NO_2, SO_2, HCHO, and CHOCHO over the Himalayan foothills: Insights from MAX-DOAS, TROPOMI, and GOME-2, Atmospheric Environment, 336, 120 746, https://doi.org/10.1016/j.atmosenv.2024.120746, 2024.

Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Friess, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, 3, 1629–1646, https://doi.org/10.5194/amt-3-1629-2010, 2010.

Rothman, L., Gordon, I., Barber, R., Dothe, H., Gamache, R., Goldman, A., Perevalov, V., Tashkun, S., and Tennyson, J.: HITEMP, the high-temperature molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 2139–2150, https://doi.org/10.1016/j.jqsrt.2010.05.001, 2010.

Rothman, L., Gordon, I., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P., Birk, M., Bizzocchi, L., Boudon, V., Brown, L., Campargue, A., Chance, K., Cohen, E., Coudert, L., Devi, V., Drouin, B., Fayt, A., Flaud, J.-M., Gamache, R., Harrison, J., Hartmann, J.-M., Hill, C., Hodges, J., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R., Li, G., Long, D., Lyulin, O., Mackie, C., Massie, S., Mikhailenko, S., Müller, H., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E., Richard, C., Smith, M., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G., Tyuterev, V., and Wagner, G.: The HITRAN2012 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 130, 4–50, https://doi.org/10.1016/j.jqsrt.2013.07.002, 2013.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, John Wiley & Sons, 2 edn., ISBN 9780471720188, 2006.

Seo, S., Kim, S.-W., Kim, K.-M., Lamsal, L. N., and Jin, H.: Reductions in NO2 concentrations in Seoul, South Korea detected from space and ground-based monitors prior to and during the COVID-19 pandemic, Environmental Research Communications, 3, 051 005, https://doi.org/10.1088/2515-7620/abed92, 2021.

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., and Burrows, J. P.: High spectral resolution ozone absorption cross-sections - Part 2: Temperature dependence, Atmospheric Measurement Techniques, 7, 625–636, https://doi.org/10.5194/amt-7-625-2014, 2014.

Seyler, A., Wittrock, F., Kattner, L., Mathieu-Üffing, B., Peters, E., Richter, A., Schmolke, S., and Burrows, J. P.: Monitoring shipping emissions in the German Bight using MAX-DOAS measurements, Atmospheric Chemistry and Physics, 17, 10 997–11 023, https://doi.org/10.5194/acp-17-10997-2017, 2017.

Silva, S. J., Heald, C. L., and Li, M.: Space-Based Constraints on Terrestrial Glyoxal Production, Journal of Geophysical Research: Atmospheres, 123, 13 583–13 594, https://doi.org/10.1029/2018JD029311, 2018.

Sindelarova, K., Markova, J., Simpson, D., Huszar, P., Karlicky, J., Darras, S., and Granier, C.: High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling, Earth System Science Data, 14, 251–270, https://doi.org/10.5194/essd-14-251-2022, 2022.

Sinreich, R., Merten, A., Molina, L., and Volkamer, R.: Parameterizing radiative transfer to convert MAX-DOAS dSCDs into near-surface box-averaged mixing ratios, Atmospheric Measurement Techniques, 6, 1521–1532, https://doi.org/10.5194/amt-6-1521-2013, 2013.

Soulie, A., Granier, C., Darras, S., Zilbermann, N., Doumbia, T., Guevara, M., Jalkanen, J.-P., Keita, S., Liousse, C., Crippa, M., Guizzardi, D., Hoesly, R., and Smith, S. J.: Global anthropogenic emissions (CAMS-GLOB-ANT) for the Copernicus Atmosphere Monitoring Service simulations of air quality forecasts and reanalyses, Earth System Science Data, 16, 2261–2279, https://doi.org/10.5194/essd-16-2261-2024, 2024.

Stavrakou, T., Müller, J.-F., Smedt, I. D., Roozendael, M. V., Kanakidou, M., Vrekoussis, M., Wittrock, F., Richter, A., and Burrows, J. P.: The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, Atmospheric Chemistry and Physics, 9, 8431–8446, https://doi.org/10.5194/acp-9-8431-2009, 2009a.

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L., and Guenther, A.: Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, Atmospheric Chemistry and Physics, 9, 1037–1060, https://doi.org/10.5194/acp-9-1037-2009, 2009b.

Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O_2-O_2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure, Physical Chemistry Chemical Physics, 15, 15 371–15 381, https://doi.org/10.1039/c3cp50968k, 2013.

Turner, M. C., Jerrett, M., Pope, C. A., Krewski, D., Gapstur, S. M., Diver, W. R., Beckerman, B. S., Marshall, J. D., Su, J., Crouse, D. L., and Burnett, R. T.: Long-Term Ozone Exposure and Mortality in a Large Prospective Study, American Journal of Respiratory and Critical Care Medicine, 193, 1134–1142, https://doi.org/10.1164/rccm.201508-1633oc, 2016.

Vandaele, A. C., Hermans, C., Simon, P. C., Roozendael, M. V., Guilmot, J. M., Carleer, M., and Colin, R.: Fourier transform measurement of NO_2 absorption cross-section in the visible range at room temperature, Journal of Atmospheric Chemistry, 25, 289–305, https://doi.org/10.1007/bf00053797, 1996.

Volkamer, R., Spietz, P., Burrows, J., and Platt, U.: High-resolution absorption cross-section of glyoxal in the UV–vis and IR spectral ranges, Journal of Photochemistry and Photobiology A: Chemistry, 172, 35–46, https://doi.org/10.1016/j.jphotochem.2004.11.011, 2005.

Volkamer, R., Martini, F. S., Molina, L. T., Salcedo, D., Jimenez, J. L., and Molina, M. J.: A missing sink for gas-phase glyoxal in Mexico City: Formation of secondary organic aerosol, Geophysical Research Letters, 34, https://doi.org/10.1029/2007gl030752, 2007.

Vrekoussis, M., Wittrock, F., Richter, A., and Burrows, J. P.: GOME-2 observations of oxygenated VOCs: what can we learn from the ratio glyoxal to formaldehyde on a global scale?, Atmospheric Chemistry and Physics, 10, 10 145–10 160, https://doi.org/10.5194/acp-10-10145-2010, 2010.

Wang, H., Li, Y., Liu, Y., Lu, X., Zhang, Y., Fan, Q., Shen, C., Lai, S., Zhou, Y., Zhang, T., and Yue, D.: Underappreciated contributions of biogenic volatile organic compounds from urban green spaces to ozone pollution, Atmospheric Chemistry and Physics, 25, 5233–5250, https://doi.org/10.5194/acp-25-5233-2025, 2025.

Welch, B. L.: The generalization of 'Student's' problem when several different population variances are involved, Biometrika, 34, 28–35, https://doi.org/10.1093/biomet/34.1-2.28, 1947.

Welch, B. L.: On the comparison of several mean values: an alternative approach, Biometrika, 38, 330, https://doi.org/10.2307/2332579, 1951.

Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., and Burrows, J. P.: MAX-DOAS measurements of atmospheric trace gases in Ny-Ålesund - Radiative transfer studies and their application, Atmospheric Chemistry and Physics, 4, 955–966, https://doi.org/10.5194/acp-4-955-2004, 2004.

Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., Volkamer, R., Beirle, S., Platt, U., and Wagner, T.: Simultaneous global observations of glyoxal and formaldehyde from space, Geophysical Research Letters, 33, https://doi.org/10.1029/2006gl026310, 2006.

Xing, C., Liu, C., Hu, Q., Fu, Q., Lin, H., Wang, S., Su, W., Wang, W., Javed, Z., and Liu, J.: Identifying the wintertime sources of volatile organic compounds (VOCs) from MAX-DOAS measured formaldehyde and glyoxal in Chongqing, southwest China, Science of The Total Environment, 715, 136 258, https://doi.org/10.1016/j.scitotenv.2019.136258, 2020.

Xing, J., Hong, Q., Yang, B., Xing, C., Yang, S., Mao, M., Su, W., Chen, Y., and Zhang, C.: Source identification of volatile organic compounds precursors from glyoxal and secondary formaldehyde utilizing MAX-DOAS observations in guangzhou, China, Atmospheric Research, 325, 108 256, https://doi.org/10.1016/j.atmosres.2025.108256, 2025.

Zarzana, K. J., Min, K.-E., Washenfelder, R. A., Kaiser, J., Krawiec-Thayer, M., Peischl, J., Neuman, J. A., Nowak, J. B., Wagner, N. L., Dubè, W. P., St. Clair, J. M., Wolfe, G. M., Hanisco, T. F., Keutsch, F. N., Ryerson, T. B., and Brown, S. S.: Emissions of Glyoxal and Other Carbonyl Compounds from Agricultural Biomass Burning Plumes Sampled by Aircraft, Environmental Science & Technology, 51, 11 761–11 770, https://doi.org/10.1021/acs.est.7b03517, 2017.

Zarzana, K. J., Selimovic, V., Koss, A. R., Sekimoto, K., Coggon, M. M., Yuan, B., Dubé, W. P., Yokelson, R. J., Warneke, C., de Gouw, J. A., Roberts, J. M., and Brown, S. S.: Primary emissions of glyoxal and methylglyoxal from laboratory measurements of open biomass burning, Atmospheric Chemistry and Physics, 18, 15 451–15 470, https://doi.org/10.5194/acp-18-15451-2018, 2018.

Zhang, B., Qiao, L., Han, H., Xie, W., and Li, L.: Variations in VOCs Emissions and Their O_3 and SOA Formation Potential among Different Ages of Plant Foliage, Toxics, 11, 645, https://doi.org/10.3390/toxics11080645, 2023.

Zheng, G., Kuang, C., Uin, J., Watson, T., and Wang, J.: Large contribution of organics to condensational growth and formation of cloud condensation nuclei (CCN) in the remote marine boundary layer, Atmospheric Chemistry and Physics, 20, 12 515–12 525, https://doi.org/10.5194/acp-20-12515-2020, 2020.

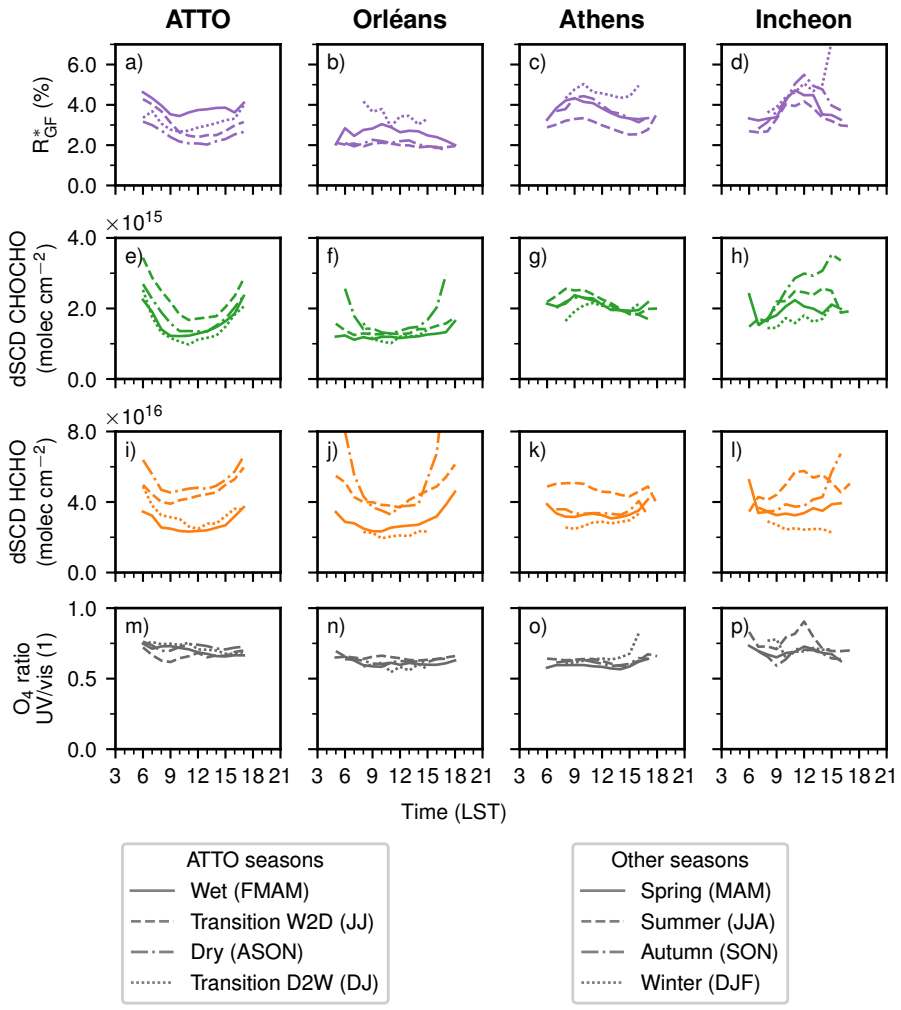

Figure A1. Diurnal cycles of R^*_{GF} (top row), CHOCHO dSCD (upper centre row), HCHO dSCD (lower centre row), and O_4 ratio (bottom row) filtered by season for ATTO, Orléans, Athens, and Incheon relative to local solar time (LST). Note that the months for ATTO are grouped differently to account for wet and dry season. Panels e) and i) are self-created based on Donner (2024).

Appendix

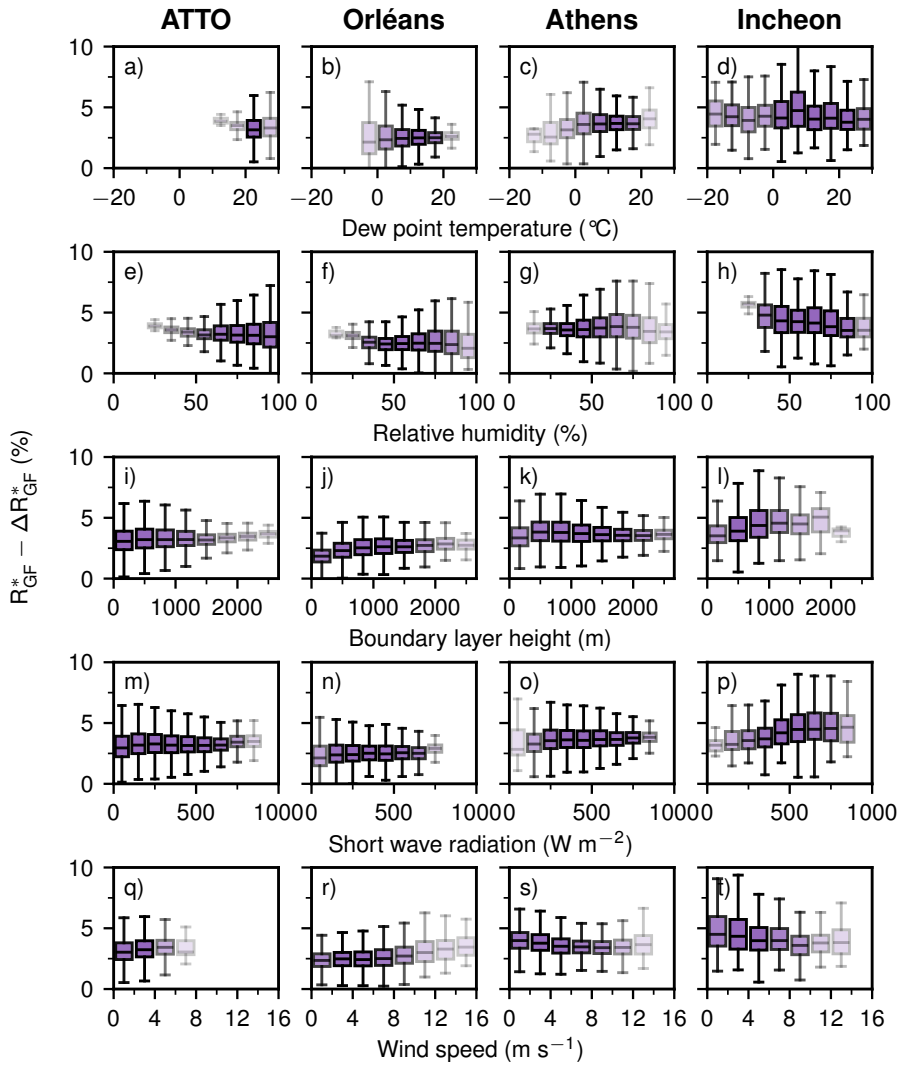

Figure A2. Temperature-normalised R^*_{GF} as a function of binned meteorological variables from ERA5. Within each box, the horizontal line indicates the median and the box spans the IQR; whiskers extend to 1.5 IQR. Box transparency scales with the number of contributing measurements, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval.

Supplement

Table A1. Fit settings for NO_2 vis for Orléans, Athens, and Incheon.

Fit parameter		Selection/source
Spectral range		425 nm – 490 nm
Polynomial degree		6
Wavelength calibration		Solar atlas from Chance and Kurucz (2010)
Reference		Closest two zenith-sky spectra in time interpolated to time of measurement
Cross-section	Temperature	Data source
O_3	223 K	Serdyuchenko et al. (2014) with I_0 correction
NO_2	294 K, 220 K	Vandaele et al. (1996) with I_0 correction
O_4	293 K	Thalman and Volkamer (2013)
H_2O	296 K	Rothman et al. (2010)
Ring	-	Computed by QDOAS, Danckaert et al. (2017)

Table A2. Fit settings for NO_2 UV for Orléans, Athens, and Incheon.

Fit parameter		Selection/source
Spectral range		338 nm – 370 nm
Polynomial degree		6
Wavelength calibration		Solar atlas from Chance and Kurucz (2010)
Reference		Closest two zenith-sky spectra in time interpolated to time of measurement
Cross-section	Temperature	Data source
O_3	223 K, 243 K	Serdyuchenko et al. (2014) with I_0 correction
NO_2	294 K, 220 K	Vandaele et al. (1996) with I_0 correction
BrO	223 K	Fleischmann et al. (2004)
HCHO	298 K	Meller and Moortgat (2000)
O_4	293 K	Finkenzeller and Volkamer (2022)
Ring	-	Computed by QDOAS with and without Pukite Terms, Danckaert et al. (2017)

Table A3. Fit settings HCHO for Orléans, Athens, and Incheon.

Fit parameter		Selection/source
Spectral range		337.5 nm – 361 nm
Polynomial degree		6
Wavelength calibration		Solar atlas from Kurucz et al. (1984)
Reference		Closest zenith-sky spectrum in time
Cross-section	Temperature	Data source
HCHO	298 K	Meller and Moortgat (2000)
O_3	223 K	Serdyuchenko et al. (2014)
NO_2	294 K	Vandaele et al. (1996)
O_4	293 K	Thalman and Volkamer (2013)
Ring	-	Computed by QDOAS, Danckaert et al. (2017)

Table A4. Fit settings CHOCHO for Orléans, Athens, and Incheon.

Fit parameter		Selection/source
Spectral range		436 nm – 470 nm
Polynomial degree		6
Wavelength calibration		Solar atlas from Kurucz et al. (1984)
Reference		Closest two zenith-sky spectra in time interpolated to time of measurement
Cross-section	Temperature	Data source
CHOCHO	296 K	Volkamer et al. (2005)
O_3	223 K	Serdyuchenko et al. (2014)
NO_2	294 K, 220 K	Vandaele et al. (1996)
O_4	293 K	Thalman and Volkamer (2013)
H_2O	293 K	Rothman et al. (2013)
Ring	-	Computed by QDOAS, Danckaert et al. (2017)

Table A5. Station specific thresholds used for data filtering. Valid datapoints fulfil the listed criteria.

	ATTO	Orléans
RMS	< 0.001	< 0.001
Intensity (~~VIS~~vis)	~~> 5e4~~ $> 5 \times 10^4$ counts s^{-1}	~~> 5e4~~ $> 5 \times 10^4$ counts s^{-1}
Intensity (UV)	~~> 1e4~~ $> 1 \times 10^4$ counts s^{-1}	~~> 1e4~~ $> 1 \times 10^4$ counts s^{-1}
SZA	< 80 °	< 80 °
rel. Error	< 50 %	< 50 %
	Incheon	**Athens**
RMS	< 0.001	< 0.001
Intensity (~~VIS~~vis)	~~> 5e4~~ $> 5 \times 10^4$ counts s^{-1}	~~-~~ -
Intensity (UV)	~~> 3e3~~ $> 3 \times 10^3$ counts s^{-1}	~~-~~ -
SZA	< 80 °	< 80 °
rel. Error	< 50 %	< 50 %

Table A6. Statistical test results for environment differences (Welch t-test) and site-to-site differences (Welch ANOVA).

Test	Data	Aggregation	Logarithmic	Value	p value	$p < 0.05$
Welch t-test	(ATT+Orl)–(Ath+Inc)	monthly mean	yes	$t = -5.8$	8×10^{-8}	yes
Welch ANOVA	all stations	monthly mean	yes	$F = 19$	3×10^{-8}	yes

Table A7. Post-hoc Games–Howell test results for pairwise station comparisons.

Pair	Aggregation	Logarithmic	Value	p value	$p < 0.05$
ATTO–Orléans	monthly mean	yes	0.16	6.7×10^{-2}	no
ATTO–Athens	monthly mean	yes	−0.15	3.2×10^{-2}	yes
ATTO–Incheon	monthly mean	yes	−0.31	2.5×10^{-5}	yes
Orléans–Athens	monthly mean	yes	−0.32	7.4×10^{-5}	yes
Orléans–Incheon	monthly mean	yes	−0.47	1.7×10^{-7}	yes
Athens–Incheon	monthly mean	yes	−0.15	6.1×10^{-2}	no

Table A8. Welch t-test results for weekday vs. weekend differences.

Station	Data	Aggregation	Logarithmic	t statistic	p value	$p < 0.05$
ATTO	weekday–weekend	weekly mean	no	−0.3	8×10^{-1}	no
Orléans	weekday–weekend	weekly mean	no	−0.2	8×10^{-1}	no
Athens	weekday–weekend	weekly mean	no	4.4	2×10^{-5}	yes
Incheon	weekday–weekend	weekly mean	no	2.7	8×10^{-3}	yes

Table A9. Seasonal and annual median absolute uncertainties of R^*_{GF} for all stations (ATTO scaled). ATTO seasons follow tropical definitions. The values in brackets are the 25%- and 75% percentile.

		ATTO		Orléans	Athens	Incheon
Annual		0.38 (0.27–0.60) %		0.64 (0.42–1.05) %	0.58 (0.41–0.90) %	0.91 (0.68–1.19) %
	DJ (transition)	0.51 (0.35–0.75) %	DJF (winter)	1.51 (1.08–2.19) %	1.06 (0.79–1.49) %	1.21 (1.05–1.48) %
	FMAM (wet)	0.56 (0.41–0.85) %	MAM (spring)	0.84 (0.56–1.33) %	0.69 (0.49–0.98) %	0.96 (0.75–1.21) %
	JJ (transition)	0.33 (0.24–0.46) %	JJA (summer)	0.51 (0.36–0.75) %	0.42 (0.33–0.53) %	0.75 (0.56–0.97) %
	ASON (dry)	0.29 (0.23–0.40) %	SON (autumn)	0.58 (0.32–1.10) %	0.76 (0.57–1.08) %	0.90 (0.68–1.14) %

Table A10. Station specific variables. The angle of the azimuth viewing direction follows N=0° and E=90°.

	ATTO	Orléans
period	12.10.2017 –31.07.2022	03.07.2023 –01.07.2024
lowest viewing elevation	2°[1]	1°
azimuth viewing direction	50°	28°
	Incheon	**Athens**
period	06.10.2021 –15.11.2022	01.01.2021 –31.12.2023
lowest viewing elevation	3°[2]	1°
azimuth viewing direction	137.5°	232.5°

[1] The highest O_4 dSCDs occurred for 2° elevation

[2] Lower elevations are obstructed

Table A11. O_4 vis scaling factors (f) applied to ATTO data per elevation. Scaling factors for ATTO were determined by matching the slope of ATTO to the total data points per elevation without ATTO

Elevation	f	Figure
2°	0.502	Fig. 1
4°	0.533	Fig. A19
8°	0.604	Fig. A20
15°	0.685	Fig. A21
30°	0.623	Fig. A22

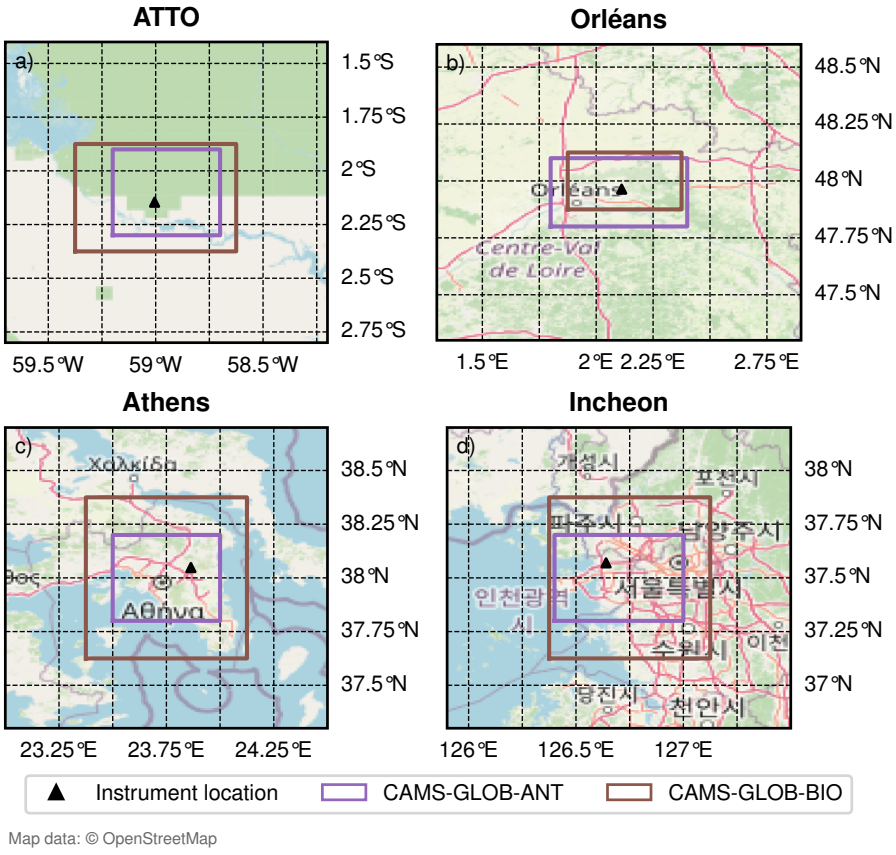

Figure A3. Maps of ~~Athens and Incheon~~ all four sites showing the considered area ~~(red square)~~ for computing the ~~CAMS~~ average anthropogenic CAMS emissions ~~for~~ (purple) and the ~~relative contributions of the different sectors, see Fig~~average biogenic CAMS emissions (brown). ~~12. Maps data: Natural Earth~~

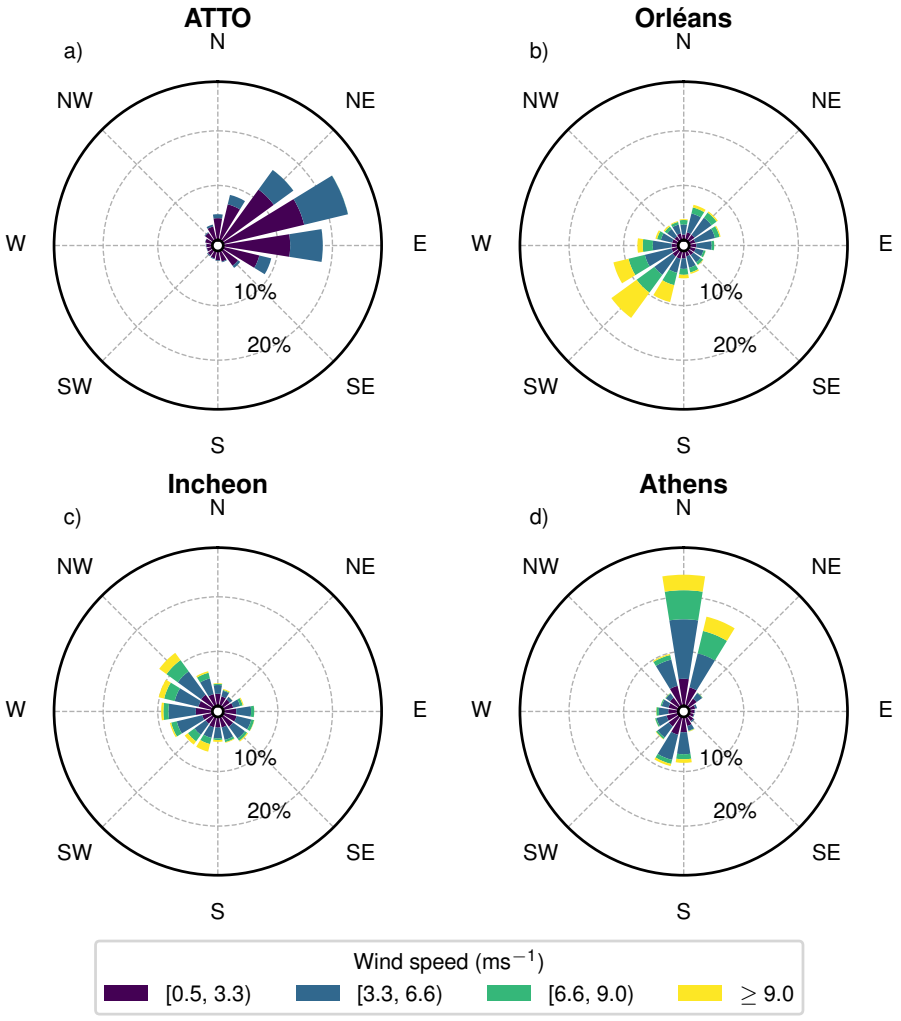

Figure A4. Wind roses showing the prevailing wind directions and wind speed in percentage during the respective measurement period for ATTO a), Orléans b), Incheon c) and Athens d). The bars show the percentage of wind direction from that direction and the colour indicates the wind speed.

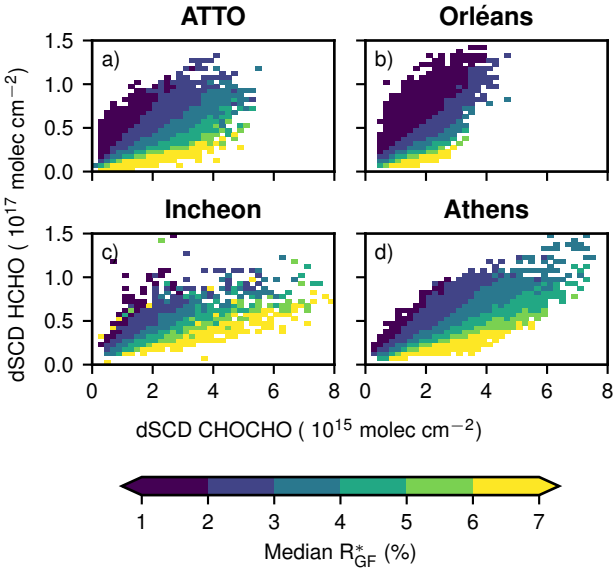

Figure A5. Median R^*_{GF} values for different bins of CHOCHO dSCDs and HCHO dSCDs.

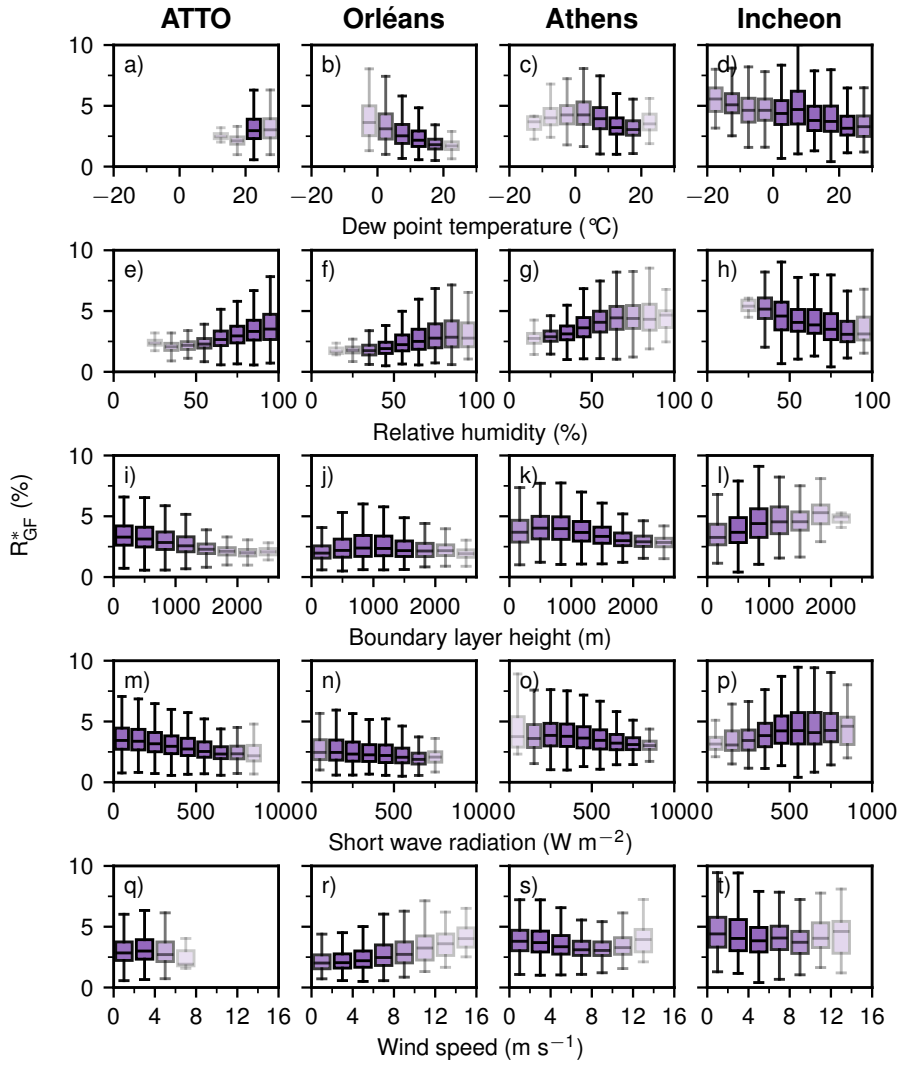

Figure A6. R^*_{GF} as a function of binned meteorological variables from ERA5. Within each box, the horizontal line indicates the median and the box spans the IQR; whiskers extend to 1.5 IQR. Box transparency scales with the number of contributing measurements, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval.

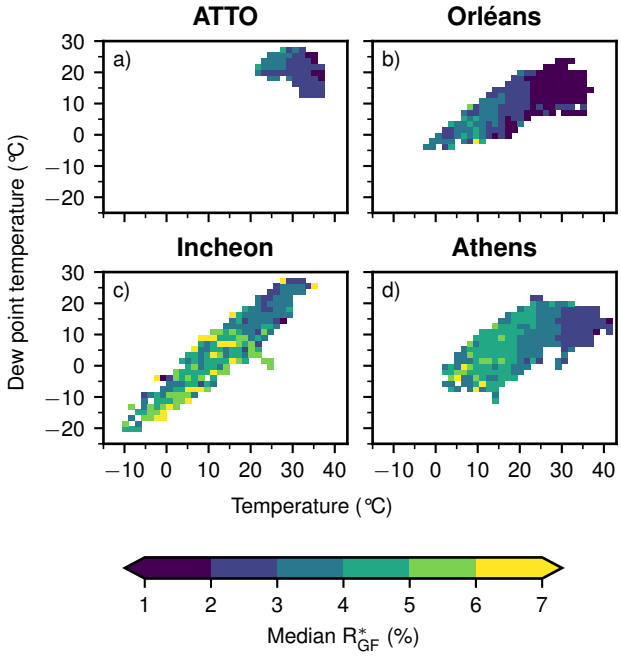

Figure A7. Median R^*_{GF} values for different bins of temperature and dew point temperature.

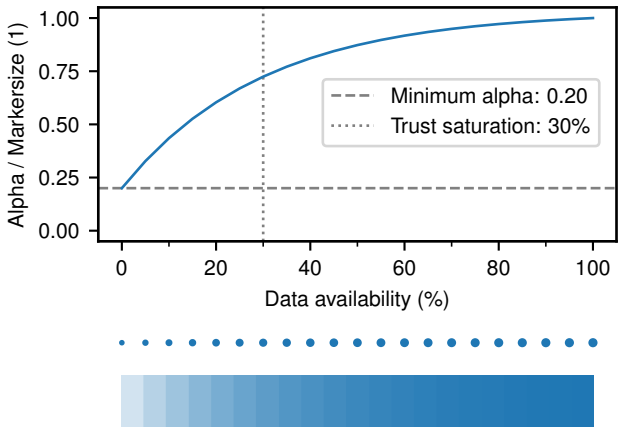

Figure A8. Applied mapping to show data availability as alpha or marker size. The data availability is defined per panel, i.e. the bin with highest number of measurements in that panel is taken as 100%. The mapping corresponds to $f(x) = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})(1 - \exp(-\frac{x}{f_0}))(1 - \exp(-\frac{1}{f_0}))^{-1}$ with $\alpha_{\min} = 0.2$, $\alpha_{\max} = 1$, $f_0 = 0.3$.

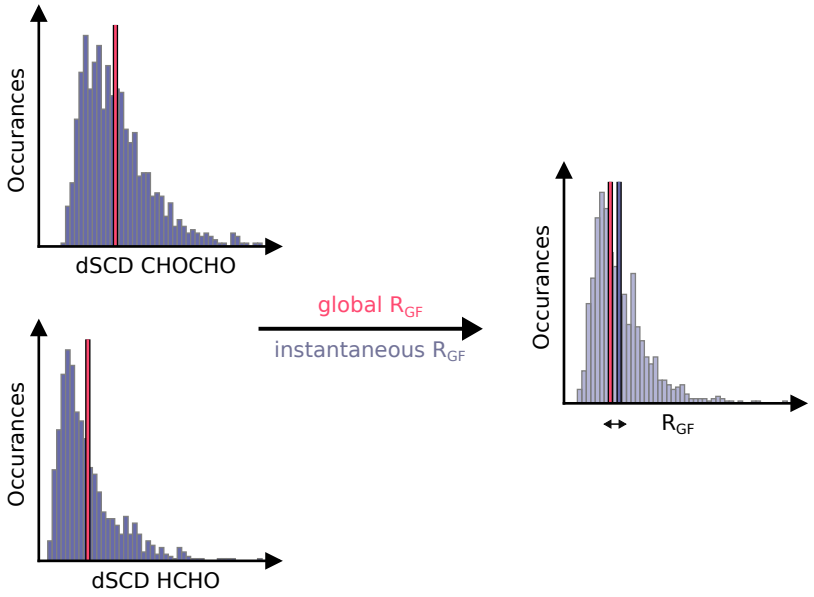

Figure A9. Visualization of the different averaging approaches using the 10~~am~~:00 local solar time bin from the ~~lowest viewing elevation from~~diurnal cycle of Orléans. The distributions of CHOCHO dSCD (top) and of HCHO dSCD (bottom) are shown on the left, with the respective mean indicated with the red vertical line. On the right, the distribution of R_{GF} and the instantaneous R_{GF} (purple vertical line) ~~is shown. The vertical red line corresponds to~~ as well as the global R_{GF} (red vertical line) is shown.

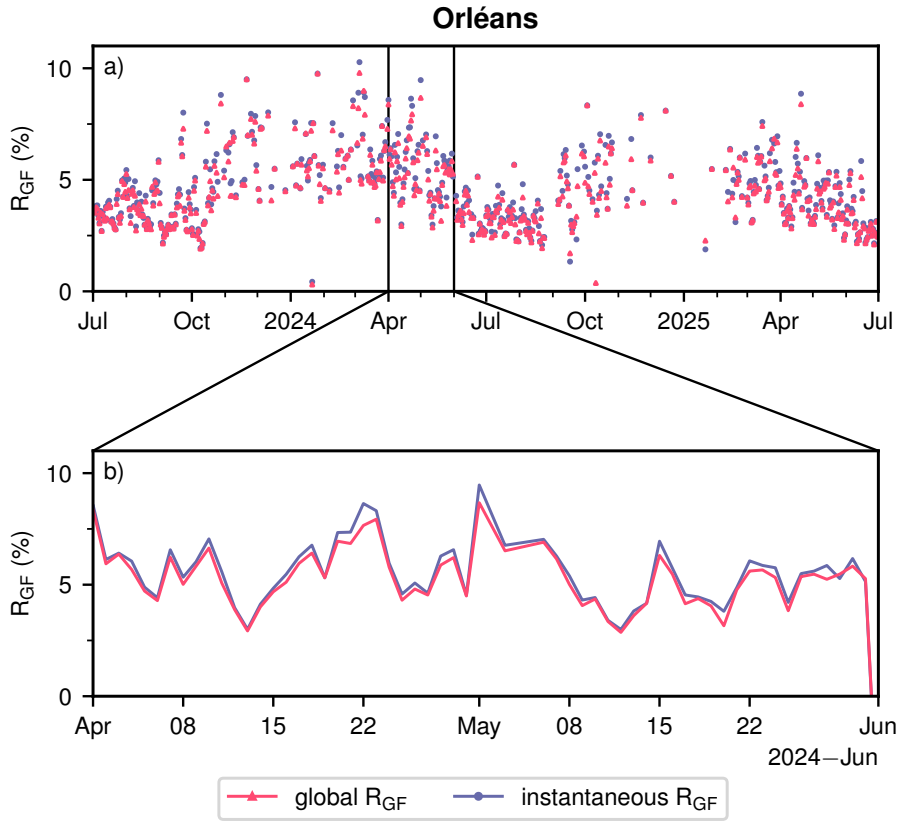

Figure A10. Timeseries of global R_{GF} and instantaneous R_{GF} for daily means at Orléans. Panel a) shows the whole measurement period. Panel b) shows a zoomed-in period.

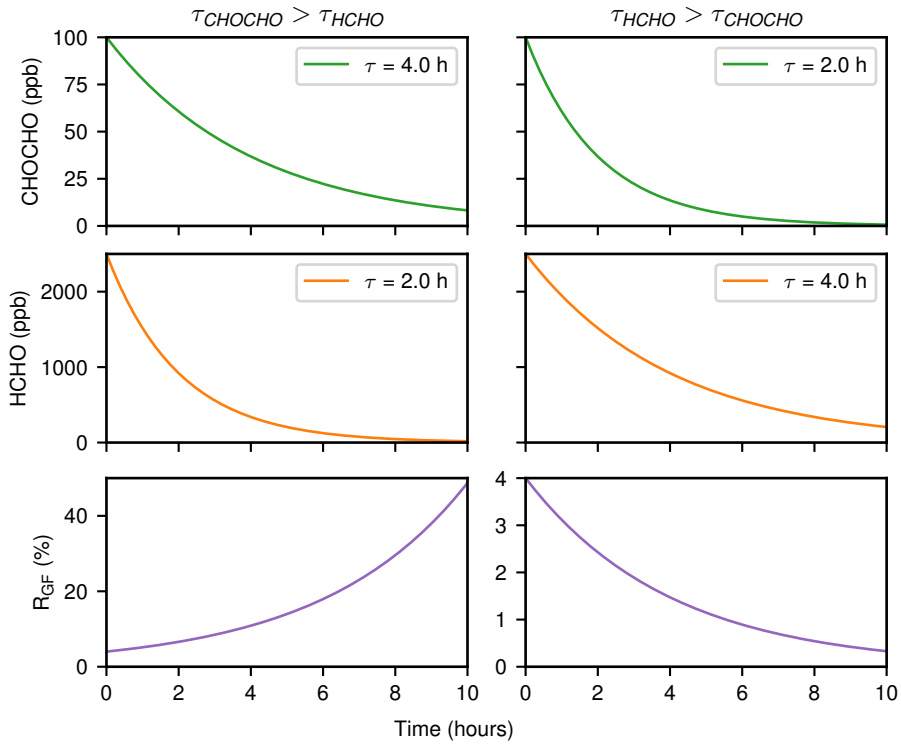

Figure A11. Simplified R_{GF} sensitivity to differences in chemical lifetime assuming an exponential decay.

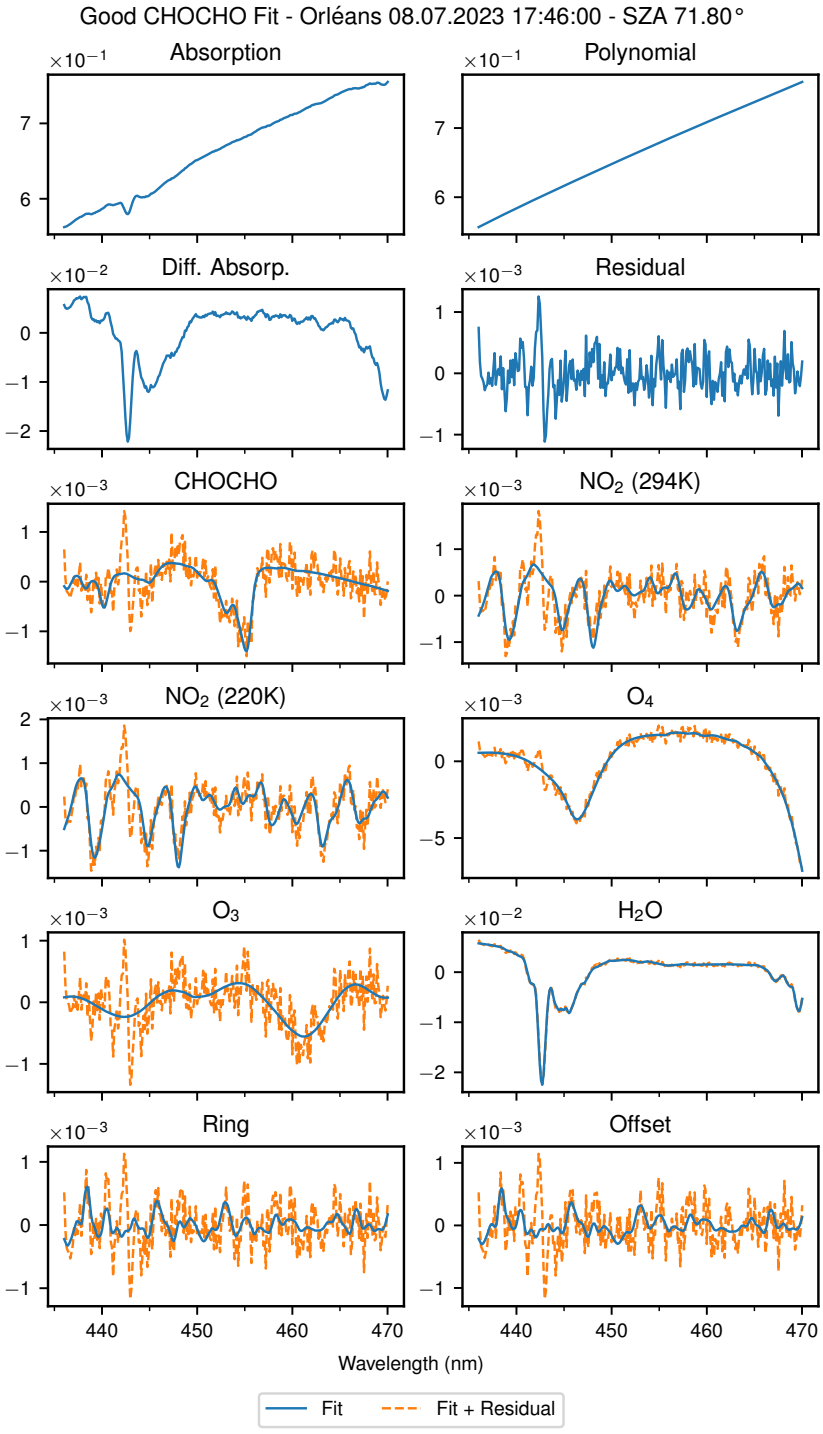

Figure A12. Fit components of a good CHOCHO fit in Orléans. The blue line indicates the fit and the dashed orange line indicates the residual added to the fit.

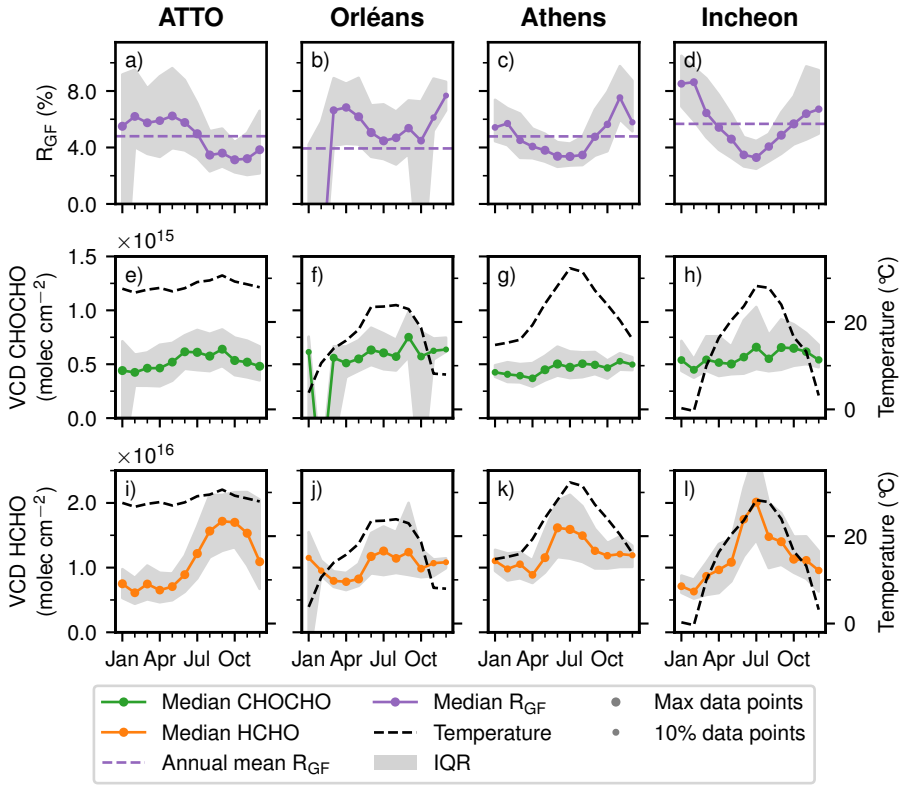

Figure A13. Seasonal cycle of R_{GF} without O_4 correction (top row), CHOCHO VCDs (centre row), HCHO VCDs (bottom row) for ATTO, Orléans, Athens, and Incheon. Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements. The VCDs are approximated by using the 30 ° viewing elevation. Panels e) and i) are are self-created based on Donner (2024).

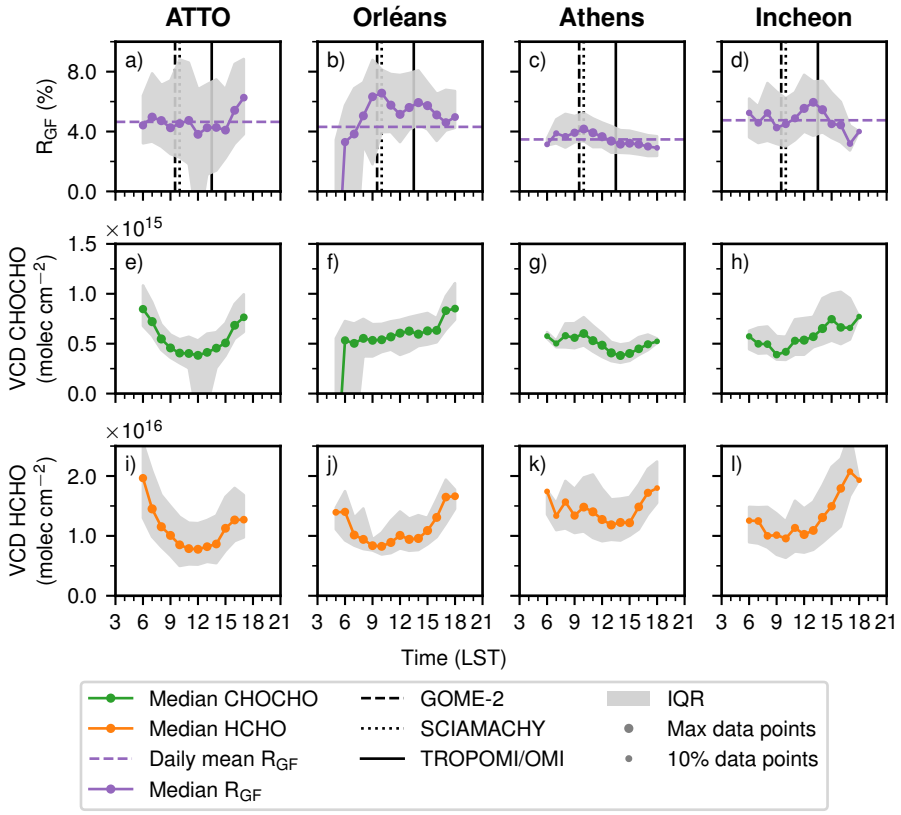

Figure A14. Diurnal cycle of R_{GF} without O_4 correction (top row), CHOCHO VCDs (centre row), HCHO VCDs (bottom row) for ATTO, Orléans, Athens, and Incheon. Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements. The VCDs are approximated by using the 30 ° viewing elevation. Panels e) and i) are are self-created based on Donner (2024).

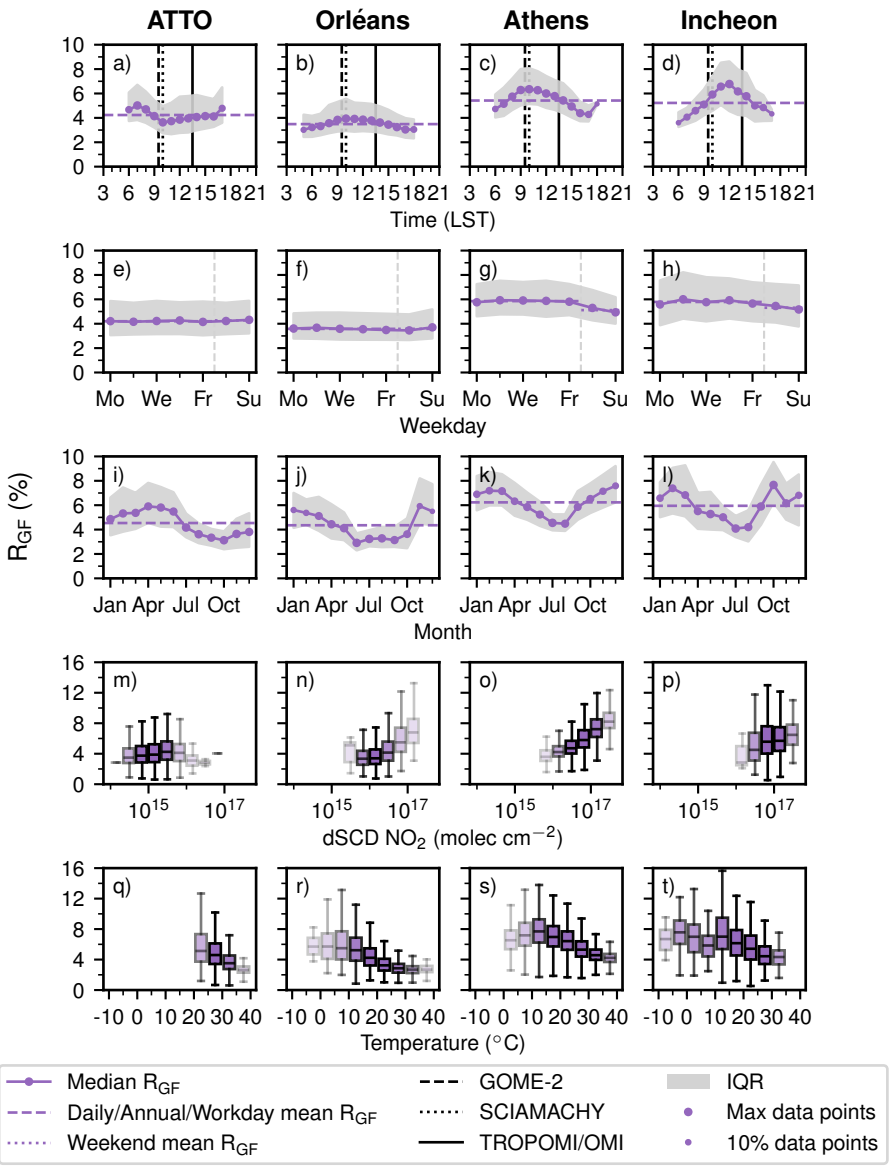

Figure A15. Overview of all main figures of the study for R_{GF} (without O_4 correction). Within each box, the horizontal line indicates the median and the box spans the IQR; whiskers extend to 1.5 IQR. Box transparency scales with the number of contributing measurements, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval.

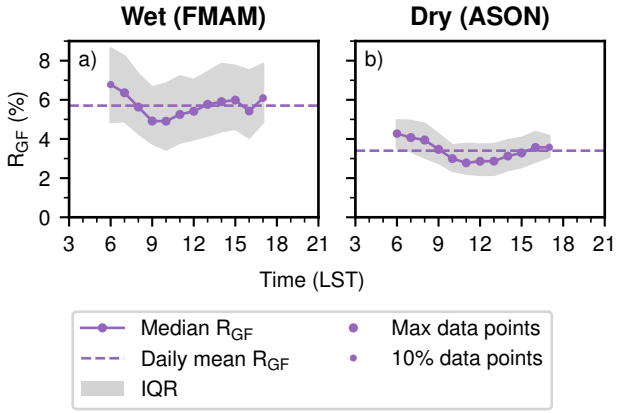

Figure A16. Diurnal cycles in the wet a) and dry b) season of R_{GF} (without O_4 correction) at ATTO. ~~To indicate the number of measurements that contribute to each data point, the size of the marker is varied. The solid line corresponds to the median value, the dashed line to the daily mean, and the gray shaded area to the IQR. Only data from the lowest viewing elevation is included.~~Marker size scales with the number of contributing observations, with smaller markers indicating fewer measurements.

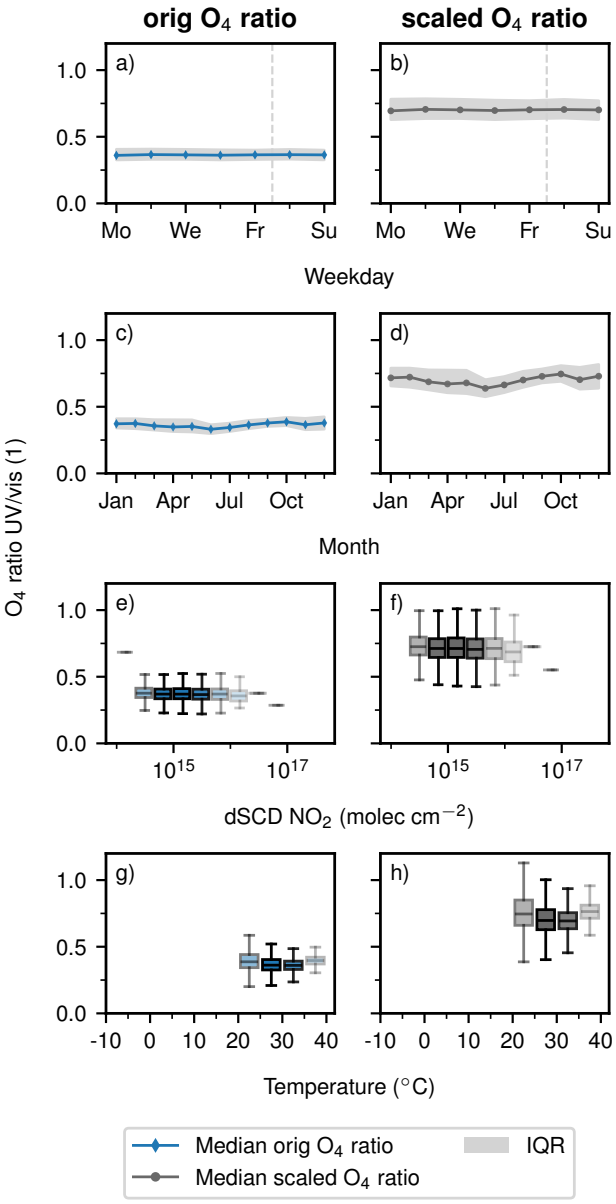

Figure A17. Overview of the behaviour of the O_4 ratio with and without scaling the O_4 vis dSCDs for ATTO. Within each box, the horizontal line indicates the median and the box spans the IQR; whiskers extend to 1.5 IQR. Box transparency scales with the number of contributing measurements, with more transparent boxes indicating fewer observations. Missing box plots indicate that no data points fall within that interval.

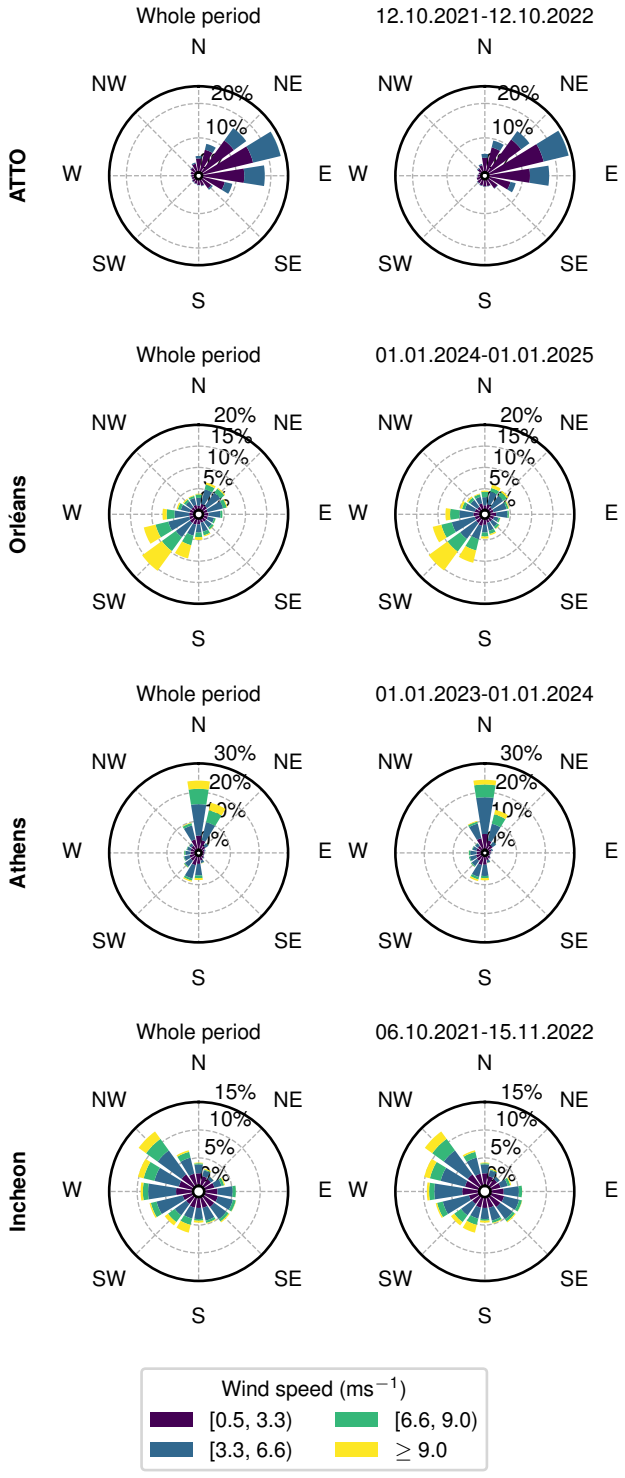

Figure A18. Wind rose of the whole measurement period and the wind rose of the subperiod that is selected for the backward trajectory runs.

Figure A19. Scatter plots showing the correlation of O_4 dSCDs from the UV wavelength range with the vis wavelength range for different datasets at 4° viewing elevation. The light grey dashed line indicates the 1-to-1 line and the gray solid line indicates a orthogonal linear fit with the specified parameters. The density of the data points is indicated by the hue, denser regions are shown in dark grey.

Figure A20. Scatter plots showing the correlation of O_4 dSCDs from the UV wavelength range with the vis wavelength range for different datasets at 8° viewing elevation. The light grey dashed line indicates the 1-to-1 line and the gray solid line indicates a orthogonal linear fit with the specified parameters. The density of the data points is indicated by the hue, denser regions are shown in dark grey.

Figure A21. Scatter plots showing the correlation of O_4 dSCDs from the UV wavelength range with the vis wavelength range for different datasets at 15° viewing elevation. The light grey dashed line indicates the 1-to-1 line and the gray solid line indicates a orthogonal linear fit with the specified parameters. The density of the data points is indicated by the hue, denser regions are shown in dark grey.

Figure A22. Scatter plots showing the correlation of O_4 dSCDs from the UV wavelength range with the vis wavelength range for different datasets at 30° viewing elevation. The light grey dashed line indicates the 1-to-1 line and the gray solid line indicates a orthogonal linear fit with the specified parameters. The density of the data points is indicated by the hue, denser regions are shown in dark grey.