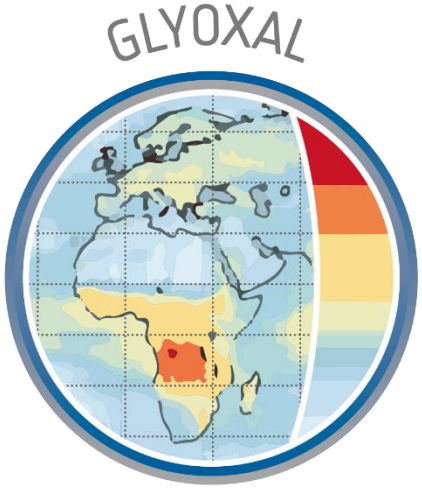
GLYoxal Retrievals from TROPOMI (GLYRETRO)

Sentinel-5p + Innovation - Theme 1: CHOCHO



Requirement Baseline Document

document number	:	S5p+I_CHOCHO_BIRA_RB
authors	:	Christophe Lerot, Trissevgeni Stavrakou, Michel Van Roozendael, Leonardo Alvarado, Andreas Richter
Distributed to	:	Christian Retscher, Marie-Helene Rio, Stefano Casadio, Daniele Gasbarra
issue	:	2.0
date	:	2019-11-01
status	:	Final

Table of Contents

Applicable Documents

[AD1] Sentinel-4 Level-2 Geophysical Algorithm Development Plan, RAL, S4-L2-RAL-GADP-2001 Issue 1.2, 2016.

[AD2] Sentinel-5 Level-2 Prototype Processor Development Requirements Specification; ESA; S5-RS-ESA-GR-0131; issue 1.6; date: 2017-01-27.

[AD3] Sentinel-5p Innovation (S5p+I) - Statement of Work - EOP-SD-SOW-2018-049, Issue 2, 20/08/2018.

[AD4] Preparing elevation data for Sentinel 5 precursor.source: KNMI, S5P-KNMI-L2-0121-TN; Issue 2.0.0, 2015-09-11.

[AD5] CAMS Forecast Data Needs for Sentinel-4 and -5 Data Processing, edited by B. Veihelmann, Issue 0.6, 15/06/2018.

[AD6] Copernicus Sentinels 4 And 5 Mission Requirements Traceability Document, EOP-SM/2413/BV-bv, Issue 2.0, 7 July 2017.

[AD7] Sentinel-5 Level-2 Prototype Processor Development Requirements Specification, ESA, S5-RS-ESA-GR-0131, issue 1.7, 2018-06-29.

[AD8] Sentinel-4 L2 Processor Component Development–Project Management Plan, DLR, S4-L2-DLR-PMP-1004, issue 2.1, 2017-05-31.

1 Introduction

The purpose of this document is to describe the past and current activities related to glyoxal (CHOCHO) retrievals from space and ground, their respective added-value and limitations, how they can contribute to this activity and what the latter will bring compared to them. It also describes the different data sets that will be used for performing the activity itself, for example to support the development of the S5p algorithm and to assess the quality of the produced data set, relying on an ensemble of defined criteria. Additionally to the overview of the related activities, this document also presents the current defined scientific and operational requirements and discusses the expected domain of compliance of the S5p product that will be developed. Related to this, a risk analysis is presented. Finally, a short description of the external input data sets necessary to this work is given as well as a description of the anticipated L2 format of the S5p product.

2 Related Activities

Glyoxal has three weak absorption bands in the visible spectral range, which makes possible the retrieval of its tropospheric columns from spectral radiance measurements. Scientific satellite glyoxal products have been already developed in the past for different instruments. The first glyoxal measurements from space have been proposed by Wittrock et al. (2006) based on SCIAMACHY/Envisat observations. Then, measurements from other nadir-viewing instruments have been exploited by different teams to derive information on glyoxal concentrations, from the GOME-2 sensors aboard the Metop platforms (Lerot et al., 2010; Vrekoussis et al., 2009) and from OMI aboard Aura (Alvarado et al., 2014; Chan Miller et al., 2014; Wang et al., 2019). All those different algorithms rely on a similar DOAS approach and most of the differences lie in the choice of the fit settings and of the auxiliary input data.

Until recently, retrievals of glyoxal tropospheric columns were considered as by-products but it is more and more recognized that the glyoxal concentration measurements can provide interesting information on volatile organic compounds (VOCs) emissions (e.g. Fu et al., 2008; Li et al., 2016; Liu et al., 2012; Stavrakou et al., 2009). Although it has similar sources as formaldehyde (HCHO), the respective production yields may differ significantly and glyoxal observations provide thus complementary information and offer other constraints than HCHO (DiGangi et al., 2012; Kaiser et al., 2015; Miller et al., 2016; Vrekoussis et al., 2010). As part of the preparation of the baseline for the future L2 Operational Processors of the Sentinel-4 and -5 missions, glyoxal has therefore been included in the list of the core products [AD1; AD2].

Glyoxal can obviously be retrieved also from ground-based DOAS instruments performing backscattered radiance measurements in the visible spectral region. Although glyoxal is often not the main driver when designing ground-based instruments, it gets more and more attention as a by-product, especially as it can be retrieved in the same spectral region as nitrogen dioxide. Nevertheless, many ground-based glyoxal column data sets are limited to punctual campaigns and the number of records long enough to facilitate a proper validation of satellite data remains limited. Fortunately, a series of interesting data sets have been recently published which could support this work. This will be addressed in the following sections.

Glyoxal is not part of the initial list of the core products derived from the TROPOMI/Sentinel-5p instrument. Since the Sentinel-5p innovation project aims at further exploiting the capability of the TROPOMI instrument, CHOCHO is logically one of the themes covered by this initiative [AD3]. In that sense, the objective and corresponding added value of this work is clear: to have at the end of the project a reliable scientific TROPOMI glyoxal product, which can be distributed to the user community and could be easily added to the current list of the operational products. Despite this development relies on the experience gained with heritage sensors, a series of aspects require to be specifically addressed to have an optimal product:

- Owing to its low optical depth, glyoxal is very sensitive to the choice of the DOAS settings. Small differences in the instrumental characteristics and their subsequent impact on the spectra may imply different optimal settings from a sensor to another. Also, the instrumental features may require some adjustments of the background correction procedure.
- Auxiliary input data need also to be adapted as a function of the instrument, especially as regards of the spatial resolution. For example, the current a priori profile databases and surface albedo climatology have spatial resolution relatively coarse compared to the TROPOMI ground pixel size and the product would benefit from developments on those aspects. As part of this work, a priori profiles will be produced with the CTM MAGRITTE at an improved resolution of 1°x1°, instead of 2°x2.5° as used in previous applications. Similarly, there are on-going efforts to produce reflectivity databases with enhanced resolution compared to current ones. Also the theme 5 of the S5p+I project aims at generating among others a TROPOMI BRDF product, which might be tested in this theme for possibly improving the air mass factors calculations.
- The number of independent ground-based glyoxal data sets is small and they are in general generated on a best-effort basis by scientific teams, like the satellite glyoxal data sets. Therefore, validation activities of glyoxal measurements from space have been limited so far. This activity offers a framework to collect independent data from third parties and to foster such activities.

3 Assessment of Existing Items

The glyoxal DOAS retrieval algorithm relies on a two-step approach: a spectral fit is first carried out to derive a glyoxal slant column density (i.e. a concentration integrated along the atmospheric effective light path), which is then converted into a vertical column density by means of a radiative transfer modelling. Those two steps require a series of items, which are listed and described below.

3.1 Spectroscopic data

The DOAS spectral fit consists in adjusting at best the optical depths of atmospheric species absorbing in the same spectral region as glyoxal, i.e. around 420-470 nm. Therefore, this procedure requires the use of pre-measured absorbing cross-sections, possibly at various temperatures if their absorption strength depends on the atmospheric temperature. Because the glyoxal optical depth is typically one order of magnitude less than typical optical depths of other absorbers in the visible spectral range, it is prone to spectral interferences and small inaccuracies in the used cross-sections may influence significantly the glyoxal slant column fit. New absorption cross-sections are regularly released and those relevant for this study need to be therefore evaluated. Table 1 lists the data sets that are currently used in the glyoxal baseline algorithm.

Table 1 : List and references of absorbing cross-sections currently used in the glyoxal DOAS fit.

Species	Reference
Glyoxal	Volkamer et al., 2005
Ozone	Gorshelev et al., 2014; Serdyuchenko et al., 2014 @ 223K
H_2O (vapor)	Rothman et al., 2013 @ 293K
NO_2	Vandaele et al., 1998 @ 220 and 296 K
O_4 (O_2-O_2)	Thalman and Volkamer, 2013 @ 293K
H_2O (liquid)	Mason et al., 2016

3.2 Surface reflectivity data bases

The sensitivity to the lowermost layers of the atmosphere and to the surface is higher in the visible spectral range than in the ultraviolet. The a priori knowledge of the

surface albedo is therefore very important, and potentially an important source of error. Ideally, LER databases covering the visible spectral range and constructed at the S5p overpass time with a spatial resolution close to the TROPOMI ground pixel size should be used. Also, it would be valuable to account somehow for the anisotropy in the surface reflection either by the use of a BRDF product if available or at least by the use of directionally-dependent LER climatology's as the neglect of this effect may introduce significant errors (Lorente et al., 2018).

Such an ideal database is not available and the current baseline is to use the OMI LER climatology (Kleipool et al., 2008), which has a spatial and time resolution of 0.5° x 0.5° and one month, respectively. Although the overpass time of OMI is similar to that of S5p, the resolution of that climatology is too coarse and neglects the observation geometry dependence. There are some on-going efforts to exploit the TROPOMI data in order to construct more suited data sets, which will contribute to reduce the uncertainties related to the knowledge of the surface reflectivity (e.g. Loyola et al., 2019). As part of the theme 5 of the S5p+I program, a BRDF product will be developed which could be tested as input of this activity.

3.3 Surface elevation data bases

The sensitivity of the nadir-viewing instruments to the lowermost layer of the atmosphere also depends on the ground elevation. This aspect is easier to address than the ground reflectivity as its variability in time is generally negligible and such an information is available at much higher resolution than our needs. We use the GMTED2010 data set that has a spatial resolution up to 7.5-arc-second (Danielson and Gesch, 2011). This database is degraded at a coarser resolution (~10km), more representative of the TROPOMI measurements as described in [AD4]. Uncertainties associated to this item are minor.

3.4 Glyoxal concentration vertical distribution

An essential item required for the computation of the air mass factors and the slant to vertical column conversions is the a priori knowledge of the normalized vertical distribution of the glyoxal concentrations in the atmosphere. Uncertainties are associated to this item are large for a series of reasons:

- The knowledge of the glyoxal production and destruction mechanisms are far from being well understood and consequently modelled. Only a limited number of chemical transport models (CTM) provide glyoxal fields as output and there are large uncertainties associated to them. The accuracy of such 3D glyoxal

fields depends on many additional input parameters, including emission inventories, which have their own limitations.

- Like for the surface albedo databases, when available, CTM provide glyoxal fields at poor spatial resolution compared to the observations, at best 1°x1°. It is clear that this type of resolution cannot represent realistically the glyoxal vertical distribution when the latter is produced by punctual sources. It has been shown by Heckel et al. (2011) that such a mismatch in resolution can lead to significant errors in NO_2 tropospheric column retrievals, and the same could hold for glyoxal.

In this activity, we will use a priori profiles simulated at the S5p overpass time by the BIRA-IASB CTM MAGRITTE. This CTM is the successor of the IMAGES CTM and runs at the spatial resolution of 1°x1° (instead of 2°x2.5° previously), thereby reducing the errors due to the model resolution. Chemical and deposition mechanisms have been updated (Müller et al., 2018, 2019) and the most up-to-date emission inventories will be used for modelling the glyoxal concentration fields. Over oceans, a glyoxal signal has been identified from both space and ground-based observations, which current models cannot reproduce. Therefore, we use as a priori profile a fixed parameterized profile reproducing profiles measured with an air-borne MAX-DOAS during the TORERO campaign in the tropical Pacific Ocean (January/February 2012) (Volkamer et al., 2015).

It has to be noted that those simulations will be performed based on consolidated input data when available. For this activity, the related delay is not problematic, but this approach would not be feasible in an operational processing context, which would require forecast capability from the CTM. At the moment, neither the CTM TM5 used for the TROPOMI operational processing, nor CAMS, which will be used for the Sentinel-4 and -5 processors, are able to model glyoxal fields. This issue has been addressed in [AD5].

3.5 Existing softwares

1. QDOAS is a generic multi-sensor DOAS retrieval tool developed at BIRA-IASB, being used for development and prototyping purposes. This program, extensively validated through different campaigns, has been used worldwide and for many different DOAS applications (mainly for ground-based and satellite applications). The software is portable to Windows and Unix-based operating systems and has been developed in collaboration with S[&]T. The graphical user interface is built on the Open-Source version of the Qt toolkit, a cross-platform application framework, and Qwt libraries. QDOAS is free software distributed under the terms of the GNU General Public License.

2. BeTV: Python tool library developed at BIRA-IASB to calculate tropospheric air mass factors, vertical columns, averaging kernels and error estimates (linked with input from QDOAS and VLIDORT).

3. The BIRA-IASB MAGRITTE model calculates the distribution of 175 chemical compounds, among which 136 species undergo transport, and can be run either globally at 2◦×2.5◦ or at 1◦x1◦ resolution depending on the application, or regionally at 0.5◦ × 0.5◦ resolution. The lateral boundary conditions of the regional model are provided by the global model. In the vertical, the modelled troposphere is divided in 40 levels between the Earth's surface and the lower stratosphere. The meteorological fields are provided by ECMWF ERA-Interim analyses. Most model parameterizations, including the transport scheme, inherit from the IMAGES model (Bauwens et al., 2016; Müller and Brasseur, 1995; Stavrakou et al., 2009).

4 Associated Datasets

Validation of the TROPOMI prototype glyoxal product will be an important activity of this work. Ideally, we would need to have long-term time series of independent reference glyoxal measurements covering several seasons and at different locations of the world representative of various geophysical conditions and emission regimes. As addressed before, the use of uncertain a priori information on the glyoxal vertical distribution may lead to significant uncertainties. Having independent information on this distribution would be certainly valuable.

However, the availability of such reference data is very limited and retrievals from the ground suffer from the same limitations as from space, i.e. high sensitivity to spectral interferences and sensitivity to noise, which make this activity very challenging. To mitigate this, comparisons will also be performed with glyoxal columns retrieved from other algorithms and/or satellites as well as simulated by the CTM MAGRITTE as already described before. In addition, the model can be used as a transfer function to compare with TROPOMI independent data that are either not collocated in time, or that cannot be compared directly (e.g. in situ concentration measurements).

4.1 Independent validation data sets

We have identified a series of glyoxal reference measurement data sets that will be used in this study. Table 2 lists them as well as their time coverage and geolocation. If any additional data set is released during this study and is available to the community,

it will obviously be considered in the validation data pool. Most of those data sets rely on MAX-DOAS measurements from the ground, but also from airplanes for two of them. Such measurements provide mainly glyoxal tropospheric columns. In theory, they offer the potential to gain limited information on vertical distribution (e.g. surface concentration in addition to the tropospheric column) but in practice, only the column information is retrieved for glyoxal.

Among the listed data sets, some have been generated well before the S5p launch. They can be nevertheless useful for evaluating the CTM MAGRITTE glyoxal fields, which can then be compared to the S5p observations. Similarly, the in situ measurements can perhaps be exploited to evaluate the glyoxal vertical distribution as modelled by MAGRITTE, which may impact the TROPOMI product via the AMF calculation, as mentioned before.

Table 2 : Independent glyoxal data sets that will be used for the validation of the TROPOMI glyoxal data set.

Id	Location	Measurement Period	Type	Data Provider
Xi	Xianghe (China)	2008-present	MAX-DOAS	BIRA-IASB
Uc	Uccle (Belgium)	2017-present	MAX-DOAS	BIRA-IASB
LR	La Réunion (France)	2018-present	MAX-DOAS	BIRA-IASB
Br	Bremen (Germany)	2018-present	MAX-DOAS	IUP-Bremen
Ath	Athens (Greece)	2018-present	MAX-DOAS	IUP-Bremen
Vi	Vienna (Austria)	2018-present	MAX-DOAS	IUP-Bremen
Gu	Gucheng (China)	May-Dec 2018	MAX-DOAS	USTC *
Pant	Pantnagar (India)	2017-present	MAX-DOAS	CEReS * http://atmos3.cr.chiba-u.jp/skynet/

Phi	Phimai (Thailand)	2015-present	MAX-DOAS	CEReS * http://atmos3.cr.chiba-u.jp/skynet/
BBFlux	NorthWestern US	Aug-Sept 2018	AMAX-DOAS	CIRES/UC
TROP_camp	tbd	tbd	MAX-DOAS	BIRA-IASB IUP-Bremen Others?
CINDI	Cabauw (The Netherlands)	Sept. 2016	MAX-DOAS	BIRA-IASB IUP-Bremen Others?
COPMAR	Atlantic Ocean	Oct. 2016	MAX-DOAS	IUP-Bremen
TORERO	Pacific Ocean	Jan.-Feb. 2012	AMAX-DOAS	CIRES/UC
SENEX	Southeastern US	Summer 2013	In situ	CIRES/UC
SOAS	Southeastern US	Jun.-Jul. 2013	In situ	U. Wisconsin

* USTC: University of Science and Technology of China, CEReS: Center for Environmental Remote Sensing, CIRES/UC: Cooperative Institute for Research in Environmental Sciences/University of Colorado

4.2 Satellite data sets

As it can be seen in Table 2, most of the independent data sets are located in Europe, US and Asia and are far from being representative of all regions of the world and associated geophysical conditions. As a workaround, satellite-satellite comparisons will be also carried out:

- Comparisons with the BIRA-IASB OMI glyoxal product generated with a similar algorithm as that applied to TROPOMI will provide information on possible differences related to instrumental characteristics.
- Comparison with the IUP-Bremen TROPOMI (Alvarado et al., 2014, 2019) will give insight in differences originating from the algorithms themselves.

5 Test Areas

The development of the TROPOMI glyoxal algorithm requires regular evaluations of the impact of possible changes in the retrieval baseline. A series of criteria can be used to assess the added value of such changes:

- Level of noise in the product,
- Smoothness of the retrieved glyoxal field in remote area,
- Geophysical soundness of the spatio-temporal variations of the retrieved glyoxal columns in reference regions, covering a large range of emission regimes (Tropical and mid-latitudes sites, regions dominated by fire events, biogenic or anthropogenic emissions)

The two latter criteria rely on a priori knowledge of the expected glyoxal fields, which remains inexact provided the large uncertainties associated to the glyoxal production and destruction processes. This makes the reference data sets presented in the previous section even more important, while keeping in mind that they have also their own limitations.

As mentioned before, the ideal test areas are the ground reference data sets covering several seasons in overlap with the S5p mission. We have only a few data sets matching those needs and they all located in Asia and Europe. For other regimes, only the consistency of the retrievals with other satellite products can be used.

6 Risk Analysis

With a large expertise of the consortium in DOAS retrievals from satellite and ground-based instruments for many species, including glyoxal, as well as a strong implication into the TROPOMI Mission Performance Center, the risks not to meet the general objectives of the project are very limited.

However, there are a number of minor risks, which are more related to the glyoxal specificities, and that are listed below:

- Owing to weak CHOCHO optical depth, glyoxal products, from both satellite and ground-based instruments, are characterized by a large level of noise and are sensitive to spectral interferences, which may lead to large systematic errors.

  This is a fundamental limitation. However, efficient background correction used for weak absorbers help to reduce systematic errors originating from spectral interferences but also from L1 calibration problems. Although the

signal-to-noise ratio of TROPOMI is better than that of older instruments, a significant level of noise will remain, requiring spatial/temporal averaging to extract meaningful information. The large amount of TROPOMI observations helps however to maintain higher time and spatial resolution than before.

- Another limitation is the scarcity of independent glyoxal observations in overlap with TROPOMI observations, which are needed for the validation activities.

  This has been addressed before. To complement the comparisons with reference ground-based measurements, the TROPOMI glyoxal product will be evaluated also with glyoxal fields modelled with the CTM MAGRITTE, but also with other glyoxal satellite products either from TROPOMI or from OMI.

- An artificial dependence of the glyoxal slant columns on the brightness of the scene has been identified in the past for all space sensors. This prevents using a cloud correction in the slant-to-vertical column conversion step and requires consequently a stringent cloud filtering. This may lead to coverage issues in regions contaminated by persistent clouds. Similarly, presence of aerosols may lead to higher uncertainties.

  This issue will remain, except in the unlikely case where the cross-correlation causing this artificial dependence is clearly identified and mitigated. Nevertheless, the high spatial resolution of TROPOMI and its associated large amount of data contributes to reducing such sampling issues. Scenes highly contaminated by aerosols will be flagged based on the Absorbing Aerosol index as a warning to users.

7 Scientific and Operational Requirements

7.1 Scientific Requirements

TROPOMI/S5p is, along with the future missions Sentinel-4 and -5, the space component of the European Earth Observation programme *Copernicus*. It provides crucial observations for a series of trace gases relevant to air quality and climate change monitoring. General requirements have been defined for such space observations to serve as best as possible those two themes [AD6] and requirements for glyoxal retrievals should be defined in regards to them.

Requirements have been defined for a series of key species, including particulate matter, ozone, NO_2, CO, SO_2 and HCHO. Since it has been generally considered as a second priority, no requirement on the glyoxal column uncertainty has been defined in the S4/5 MRTD [AD6]. For the requirements on horizontal resolution and revisit time, we can use those defined for the formaldehyde columns, as those two species are useful for similar applications. The spatial requirement for HCHO has been set to 5/20 km (goal/threshold) for air quality applications and relaxed to 10/50 km for climate applications. The revisit time requirement is 0.5/2 hours for air quality applications and can obviously not be met for space instruments boarded on LEO platforms such as TROPOMI. On contrary, the future Sentinel-4 instrument aboard a geostationary platform will provide a one-hour revisit time.

Unlike for TROPOMI, glyoxal is part of the initial list of core operational products for Sentinel-4 and -5. In this context, requirements on this product have also been defined [AD7, AD8] and are given in Table 3. While one single total uncertainty requirement is defined for Sentinel-4, two separate values are defined for the random and systematic components of the uncertainty in Sentinel-5. Owing to the faint glyoxal signal, its associated random uncertainty is large on individual measurements, in the range 6-10 x 10^{14} molec.cm^{-2} for TROPOMI. Thus, it generally meets the random error requirement as defined for Sentinel-5. Preliminary evaluation of the systematic errors indicate that the S5 systematic error requirement is also met in favorable conditions (clear sky conditions, no aerosol load…). This will be consolidated during this study, using among other comparisons with reference data. The Sentinel-4 requirement combining the two error components is much more difficult to meet on individual measurements. However, the random uncertainty can be reduced by averaging several observations. Provided the threshold requirement of 20 km for the spatial requirement, up to 16 TROPOMI observations may be combined to reduce the random error and thus making the Sentinel-4 requirement reachable as well. In summary, the glyoxal requirements currently defined are realistic and it is anticipated that they will be met for TROPOMI observations under favorable conditions. This will be confirmed during this study.

Table 3 : Uncertainty Requirements on glyoxal column retrievals defined for the Sentinel-4 and -5 missions.

	Uncertainty (Threshold)	Conditions
Sentinel-4	7 x 10^{14} molec.cm^{-2} or 50% (least stringent)	SZA < 60° VZA < 60° cloud fraction < 20% VCD > 5 x 10^{14} molec.cm^{-2}

Sentinel-5	Random error: < 1.5 x 10^{15} molec.cm^{-2} Systematic error: < 2.5 x 10^{14} molec.cm^{-2} or 50% (least stringent)	SZA < 70° VZA < 70°

7.2 Operational Requirements

Figure 1 shows the flow diagram of the BIRA-IASB glyoxal retrieval. It includes all different steps necessary to the CHOCHO retrieval, i.e. (1) the slant column density derived from the measured reflectance spectrum, (2) the AMF computation for the conversion of the slant column into a vertical column and (3) the background normalization required in the case of weak absorbers such as glyoxal. The interdependencies with static (purple boxes) and dynamic (light orange boxes) input data and other L2 products, such as clouds and aerosols are also represented (see section 8). The DOAS fit also uses as reference mean radiance spectra, which needs to be preprocessed. Similarly, the background correction relies on a database of slant columns retrieved in a reference sector, which needs to be regularly updated, ideally on a daily basis.

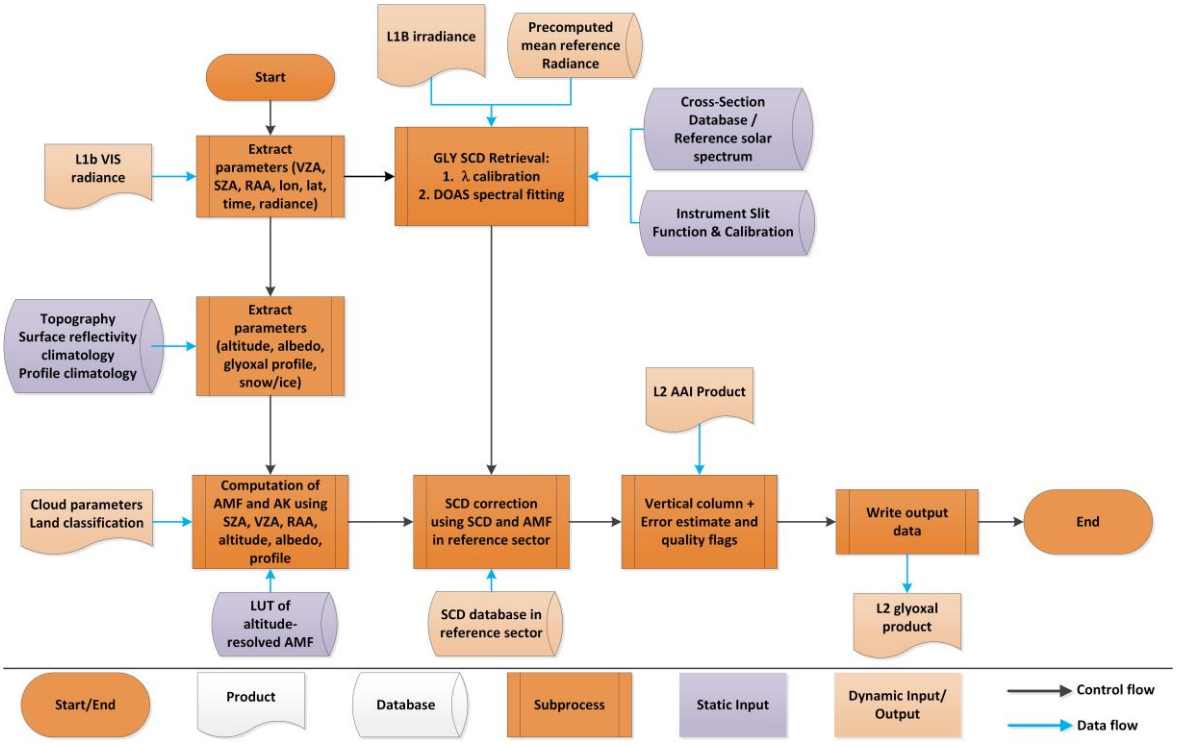


Figure 1 – General flowchart of the glyoxal DOAS retrieval algorithm.

In terms of computational approach, the different modules of the algorithm are typically executed sequentially. Using the current BIRA-IASB computing infrastructure, the process of one TROPOMI orbit requires less than 40 minutes using one single core. The DOAS fit step requires most of this time (~20 minutes), while the AMF and normalization modules need about 5 minutes each. Some additional time is in general needed in various data transfer. Those estimates are based on the initial TROPOMI spatial resolution (7 x 3.5 km²) and 20% has to be added to those numbers to account for the data rate increase after the resolution change in August 2019 (5.5 x 3.5 km²). Based on the current DOAS operational products, it can be anticipated that the size of the L2 glyoxal data will be roughly 500 Mb per orbit.

8 Input/Output Data

8.1 Input data

There is a number of input data needed for retrieving glyoxal tropospheric columns from TROPOMI observations, both static and dynamic, which are summarized in Table 4.

Main static input data sets have already been discussed in section 3. They include spectroscopic data, topography, ground reflectivity and a priori information on vertical distribution of glyoxal concentrations. In addition to the latter input, the air mass factor calculations also require altitude-dependent scattering weights representative of the instrumental sensitivity to a concentration change. Those functions depend on the observation geometry and on the scene altitude and brightness. In order to reduce computational time, scattering weights are precomputed with the Radiative Transfer model VLIDORT (vector mode: accounts for the polarization) (Spurr and Christi, 2019) and tabulated. A simple look-up through the table is thus needed during the AMF computation step. The choice of the RT model is not critical at all as it has been shown by Lorente et al. (2017) that the uncertainties on scattering weight simulations due to the model itself are very small.

In terms of dynamic input, L1 data are obviously needed. Both irradiance and radiance products covering the visible spectral range are necessary. Other dynamic inputs are required for filtering purposes. In particular, filters are currently applied to remove cloud- and/or snow-contaminated scenes. Also a land classification flag helps to discriminate land and ocean pixels for the selection of the a priori glyoxal profile. Those different flags are currently extracted from the NO2 operational product. In particular, there is in this product an estimate of the cloud fraction as retrieved directly in the NO2 spectral fit region, which is in overlap with the glyoxal fit window. This has the advantage to avoid co-registration issues between different instrumental bands, and to minimize wavelength dependence of the cloud coverage. Other cloud products

might possibly be tested. As mentioned before, no cloud correction is currently applied. If this would be changed, then cloud pressure and albedo would be needed too.

Depending on the developments that will be carried out during the project, additional parameters might be needed. For example, if a BRDF product becomes available as part of the theme 5 of S5p+I, it might be tested as additional input for a more sophisticated treatment of the surface in the glyoxal AMF computation. As mentioned before, the neglect of the surface reflectivity anisotropy may introduce some significant errors in the computed air mass factors. Differences of effective Lambertian Equivalent Reflectivity between the most eastern and western pixels of a S5p orbit may reach values up to 0.04 in the visible spectral range, which can propagate into AMF errors up to 25%. Therefore, a directional LER climatology is highly required. Uncertainties on LER values in the visible range impact significantly the glyoxal product and should be as small as possible. 0.02 is often used as a typical LER uncertainty and leads to AMF errors in the range 10-20%. If possible, this uncertainty should not be larger than that value. A more sophisticated approach would be to use as input an effective LER value computed from the BRDF parameters retrieved for every TROPOMI observation. For this, the effective LER value should be provided by the BRDF product at a wavelength close to 448 nm, also with an uncertainty close to 0.02. The gain in accuracy of this approach compared to the simple DLER climatology is likely limited. The glyoxal product would benefit mostly in case of sudden changes of the surface characteristics.

Table 4 : List of static and dynamic inputs required for generating the TROPOMI glyoxal product.

Parameter	Physical unit	Source
Static Inputs		
High-resolution solar spectrum	mol s^{-1} m^{-2} nm^{-1}	Chance and Kurucz (2010)
Absorption O_3 cross-sections at 243 K	$cm^2molec.^{-1}$	Serdyuchenko et al. (2014)
Absorption NO_2 cross-section at 220 and 296 K	$cm^2molec.^{-1}$	Vandaele et al. (1998)
Absorption glyoxal cross-section at 296K	$cm^2molec.^{-1}$	Volkamer et al. (2005)
Absorption water vapor cross-section at 293K	$cm^2molec.^{-1}$	Rothman al. (2012)
Absorption O_4 cross-section at 293K	$cm^5molec.^{-2}$	Thalman et al. (2013)
Liquid water absorption cross-section	m^{-1}	Mason et al. (2016)
Ring cross-section	---	Generated internally

Surface Albedo in the fitting window	---	Kleipool et al. (2008) Other options possible (see section 3)
Surface altitude	m	GMTED 2010
A-priori CHOCHO vertical profile shapes	vmr	CTM MAGRITTE TORERO profile over oceans (Volkamer et al., [2015])
Look-up table of altitude-resolved AMFs	---	Generated internally with VLIDORT
Dynamic Inputs		
Radiance	mol s^{-1} m^{-2} nm^{-1} sr^{-1}	L1b radiance product (Band 4)
Irradiance	mol s^{-1} m^{-2} nm^{-1}	L1b irradiance product (UVN)
Geolocation	degrees	L1b radiance product (Band 4)
Cloud fraction (Cloud pressure)	--- (hPa)	L2 operational NO_2 product
Snow/ice flag	---	L2 operational NO_2 product
Land classification flag	---	L2 operational NO_2 product

8.2 Output data

For the output format of the L2 files, we propose to follow as much as possible the current conventions of the operational products, in terms of filename and content structure as well as variable name nomenclature and units. Thus, the output files will be in NetCDF, with the convention CF.

A typical L2 glyoxal filename would be structured as:

S5P_OFFL_L2__CHOCHO____20180904T054104_20180904T072234_04623_01_010000_20190315.nc

where the two first time stamps in yellow correspond to start and end of the orbit, the orbit number is in green, collection and product versions are in grey, and the creation date is in cyan. Following the operational products, all variables provided in the L2 glyoxal files will be organized in the structure sketched as in Figure 2.

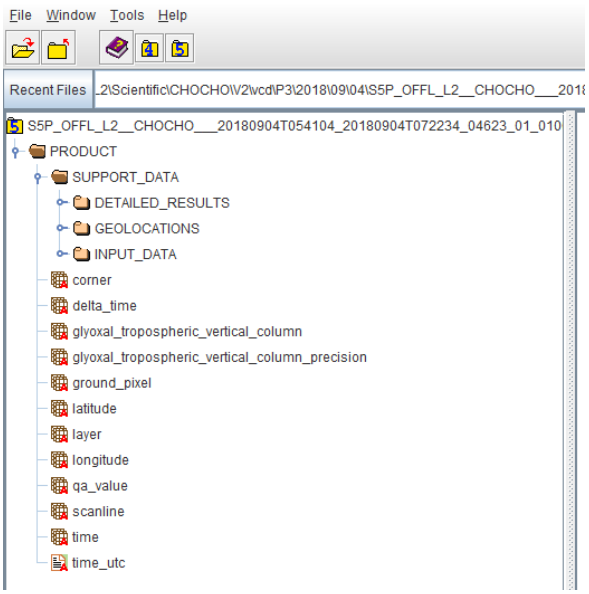


Figure 2 : L2 glyoxal file structure

Table 5 gives a preliminary list of variables expected to be provided in the TROPOMI L2 glyoxal column product. This list will be consolidated and adapted depending on the algorithm evolution.

Table 5 : Preliminary list of variables that will be included in the TROPOMI glyoxal product. The second column indicates the GROUP of the file content structure in which the variable will be stored.

Name	Group	Unit	Description / long name
time	PRODUCT	s	Reference time of the measurements.
scanline	PRODUCT	-	Coordinate variable defining the indices along track.
delta_time	PRODUCT	ms [since time]	Time difference with reference time for each scanline.
ground_pixel	PRODUCT	-	Coordinate variable defining the indices across track.
glyoxal_tropospheric_column	PRODUCT	mol.m-2	Glyoxal tropospheric column
glyoxal_tropospheric_column_precision	PRODUCT	mol.m-2	Glyoxal tropospheric column random error
glyoxal_tropospheric_column_trueness	PRODUCT	mol.m-2	Glyoxal tropospheric column systematic error
glyoxal_tropospheric_column_kernel_trueness	PRODUCT	mol.m-2	Glyoxal tropospheric column systematic error without smoothing error

qa_value	PRODUCT	-	Quality assurance value describing the quality of the product
latitude	PRODUCT	degree north	Latitude of the center of each ground pixel
longitude	PRODUCT	degree east	Longitude of the center of each ground pixel
glyoxal_slant_column	DETAILED_RESULTS	mol.m-2	Glyoxal slant column
glyoxal_slant_column_precision	DETAILED_RESULTS	mol.m-2	Glyoxal slant column random error
glyoxal_slant_column_trueness	DETAILED_RESULTS	mol.m-2	Glyoxal slant column systematic error
glyoxal_slant_column_correction	DETAILED_RESULTS	mol.m-2	Glyoxal slant column correction
glyoxal_clear_air_mass_factor	DETAILED_RESULTS	-	Clear air mass factor
glyoxal_air_mass_factor_trueness	DETAILED_RESULTS	-	Systematic error on the air mass factor
averaging_kernel	DETAILED_RESULTS	-	Averaging kernel
fitted_root_mean_square	DETAILED_RESULTS	-	DOAS fit residuals
cost_function	DETAILED_RESULTS	-	Reduced Chi-squared of the DOAS fit
irradiance_wavelength_shift	DETAILED_RESULTS	nm	Wavelength calibration shift results
irradiance_reference_wavelength	DETAILED_RESULTS	nm	Wavelength calibration reference wavelengths
fitted_slant_columns	DETAILED_RESULTS	Various	Slant column density from all absorbers in the fitting window
fitted_slant_columns_precision	DETAILED_RESULTS	Various	Slant column density random errors from all absorbers in the fitting window
fitted_radiance_shift	DETAILED_RESULTS	nm	Wavelength shift from the DOAS fit
fitted_radianc_stretch	DETAILED_RESULTS	-	Wavelength stretch from the DOAS fit
solar_zenith_angle	GEOLOCATIONS	degree	Zenith angle of the sun measured from the ground pixel location
solar_azimuth_angle	GEOLOCATIONS	degree	Azimuth angle of the sun measured from the ground pixel location
viewing_zenith_angle	GEOLOCATIONS	degree	Zenith angle of the satellite measured from the ground pixel location
viewing_azimuth_angle	GEOLOCATIONS	degree	Azimuth angle of the satellite measured from the ground pixel location
latitude_bounds	GEOLOCATIONS	degree north	The four latitude boundaries of each ground pixel.
longitude_bounds	GEOLOCATIONS	degree east	The four longitude boundaries of each ground pixel.
surface_pressure	INPUT_DATA	Pa	Surface pressure from the CTM adjusted for surface elevation.
surface_altitude	INPUT_DATA	m	Height of the surface averaged over the ground pixel.
snow_ice_flag	INPUT_DATA	-	Surface condition (snow/ice)

surface_classification	INPUT_DATA	-	Surface classification
surface_albedo	INPUT_DATA	-	Surface albedo
glyoxal_profile_apriori	INPUT_DATA	vmr	A priori glyoxal profile
glyoxal_profile_apriori_pressure	INPUT_DATA	Pa	Pressure grid of a priori glyoxal profile
cloud_fraction	INPUT_DATA	-	Cloud fraction
aerosol_index_340_380	INPUT_DATA	-	Aerosol absorbing index 340/380 pair

9 References

Alvarado, L. M. A., Richter, A., Vrekoussis, M., Wittrock, F., Hilboll, A., Schreier, S. F. and Burrows, J. P.: An improved glyoxal retrieval from OMI measurements, Atmos. Meas. Tech., 7(12), 4133–4150, doi:10.5194/amt-7-4133-2014, 2014.

Alvarado, L. M. A., Richter, A., Vrekoussis, M., Hilboll, A., Kalisz Hedegaard, A. B., 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, Atmos. Chem. Phys. Discuss., 1–22, doi:10.5194/acp-2019-684, 2019.

Bauwens, M., Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Wiedinmyer, C., Kaiser, J. W., Sindelarova, K. and Guenther, A.: Nine years of global hydrocarbon emissions based on source inversion of OMI formaldehyde observations, Atmos. Chem. Phys., 16(15), 10133–10158, doi:10.5194/acp-16-10133-2016, 2016.

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, Atmos. Meas. Tech., 7(11), 3891–3907, doi:10.5194/amt-7-3891-2014, 2014.

Chance, K. and Kurucz, R. L.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, J. Quant. Spectrosc. Radiat. Transf., 111(9), 1289–1295, doi:10.1016/j.jqsrt.2010.01.036, 2010.

Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010)., 2011.

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, Atmos. Chem. Phys., 12(20), 9529–9543, doi:10.5194/acp-12-9529-2012, 2012.

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, J. Geophys. Res., 113(D15), D15303, doi:10.1029/2007JD009505, 2008.

Gorshelev, V., Serdyuchenko, A., Weber, M., Chehade, W. and Burrows, J. P.: High spectral resolution ozone absorption cross-sections; Part 1: Measurements, data analysis and comparison with previous measurements around 293 K, Atmos. Meas. Tech., 7(2), 609–624, doi:10.5194/amt-7-609-2014, 2014.

Heckel, A., Kim, S.-W., Frost, G. J., Richter, A., Trainer, M. and Burrows, J. P.: Influence of low spatial resolution a priori data on tropospheric NO_2 satellite retrievals, Atmos. Meas. Tech., 4(9), 1805–1820, doi:10.5194/amt-4-1805-2011, 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, Atmos. Chem. Phys., 15(13), 7571–7583, doi:10.5194/acp-15-7571-2015, 2015.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F. and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113(D18), D18308, doi:10.1029/2008JD010290, 2008.

Lerot, C., Stavrakou, T., De Smedt, I., M̈uller, J. F. and Van Roozendael, M.: Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, Atmos. Chem. Phys., 10(24), 12059–12072, 2010.

Li, J., Mao, J., Min, K.-E., Washenfelder, R. A., Brown, S. S., Kaiser, J., Keutsch, F. N., Volkamer, R., Wolfe, G. M., Hanisco, T. F., Pollack, I. B., Ryerson, T. B., Graus, M., Gilman, J. B., Lerner, B. M., Warneke, C., de Gouw, J. A., Middlebrook, A. M., Liao, J., Welti, A., Henderson, B. H., McNeill, V. F., Hall, S. R., Ullmann, K., Donner, L. J., Paulot, F. and Horowitz, L. W.: Observational constraints on glyoxal production from isoprene oxidation and its contribution to organic aerosol over the Southeast United States, J. Geophys. Res. Atmos., 121(16), 9849–9861, doi:10.1002/2016JD025331, 2016.

Liu, Z., Wang, Y., Vrekoussis, M., Richter, A., Wittrock, F., Burrows, J. P., Shao, M., Chang, C.-C., Liu, S.-C., Wang, H. and Chen, C.: Exploring the missing source of glyoxal (CHOCHO) over China, Geophys. Res. Lett., 39(10), L10812, doi:10.1029/2012GL051645, 2012.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J. T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J. and Krol, M.: Structural uncertainty in air mass factor calculation for NO2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10(3), 759–782, doi:10.5194/amt-10-759-2017, 2017.

Lorente, A., Folkert Boersma, K., Stammes, P., Gijsbert Tilstra, L., Richter, A., Yu, H.,

Kharbouche, S. and Muller, J. P.: The importance of surface reflectance anisotropy for cloud and NO2 retrievals from GOME-2 and OMI, Atmos. Meas. Tech., 11(7), 4509–4529, doi:10.5194/amt-11-4509-2018, 2018.

Loyola, D. G., Xu, J., Heue, K.-P. and Zimmer, W.: Applying FP_ILM to the retrieval of geometry-dependent effective Lambertian equivalent reflectivity (GE_LER) to account for BRDF effects on UVN satellite measurements of trace gases, clouds and aerosols, Atmos. Meas. Tech. Discuss., 1–28, doi:10.5194/amt-2019-37, 2019.

Mason, J. D., Cone, M. T. and Fry, E. S.: Ultraviolet (250–550 nm) absorption spectrum of pure water, Appl. Opt., 55(25), 7163, doi:10.1364/AO.55.007163, 2016.

Miller, C. C., Jacob, D. J., Marais, E. A., Yu, K., Travis, K. R., Kim, P. S., Fisher, J. A., Zhu, L., Wolfe, G. M., Keutsch, F. N., Kaiser, J., Min, K.-E., Brown, S. S., Washenfelder, R. A., González Abad, 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, Atmos. Chem. Phys. Discuss., 1–25, doi:10.5194/acp-2016-1042, 2016.

Müller, J.-F. and Brasseur, G.: IMAGES: A three-dimensional chemical transport model of the global troposphere, J. Geophys. Res., 100(D8), 16445, doi:10.1029/94JD03254, 1995.

Müller, J.-F., Stavrakou, T., Bauwens, M., Compernolle, S. and Peeters, J.: Chemistry and deposition in the Model of Atmospheric composition at Global and Regional scales using Inversion Techniques for Trace gas Emissions (MAGRITTE v1.0). Part B. Dry deposition, Geosci. Model Dev. Discuss., 1–49, doi:10.5194/gmd-2018-317, 2018.

Müller, J.-F., Stavrakou, T. and Peeters, J.: Chemistry and deposition in the Model of Atmospheric composition at Global and Regional scales using Inversion Techniques for Trace gas Emissions (MAGRITTE v1.1) – Part 1: Chemical mechanism, Geosci. Model Dev., 12(6), 2307–2356, doi:10.5194/gmd-12-2307-2019, 2019.

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

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W. and Burrows, J. P.: High spectral resolution ozone absorption cross-sections – Part 2: Temperature dependence, Atmos. Meas. Tech., 7(2), 625–636, doi:10.5194/amt-7-625-2014, 2014.

Spurr, R. and Christi, M.: The LIDORT and VLIDORT Linearized Scalar and Vector Discrete Ordinate Radiative Transfer Models: Updates in the Last 10 Years, pp. 1–62, Springer, Cham.,

2019.

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., 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, Atmos. Chem. Phys., 9(21), 8431–8446 [online] Available from: http://www.atmos-chem-phys.net/9/8431/2009/, 2009.

Thalman, R. and Volkamer, R.: Temperature dependent absorption cross-sections of O2–O2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure, Phys. Chem. Chem. Phys., 15(37), 15371, doi:10.1039/c3cp50968k, 2013.

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Mérienne, M. F., Jenouvrier, A. and Coquart, B.: Measurements of the NO2 absorption cross-section from 42 000 cm−1 to 10 000 cm−1 (238–1000 nm) at 220 K and 294 K, J. Quant. Spectrosc. Radiat. Transf., 59(3–5), 171–184, doi:10.1016/S0022-4073(97)00168-4, 1998.

Volkamer, R., Spietz, P., Burrows, J. and Platt, U.: High-resolution absorption cross-section of glyoxal in the UV—vis and IR spectral ranges, J. Photochem. Photobiol. A Chem., 172(1), 35—46 [online] Available from: http://www.colorado.edu/chemistry/volkamer/publications/articles/Volkamer etal (2005) HR cross section glyoxal.pdf, 2005.

Volkamer, R., Baidar, S., Campos, T. L., Coburn, S., DiGangi, J. P., Dix, B., Eloranta, E. W., Koenig, T. K., Morley, B., Ortega, I., Pierce, B. R., Reeves, M., Sinreich, R., Wang, S., Zondlo, M. A. and Romashkin, P. A.: Aircraft measurements of BrO, IO, glyoxal, NO2, H2O, O2–O2 and aerosol extinction profiles in the t, Atmos. Meas. Tech., 8(5), 2121–2148, doi:10.5194/amt-8-2121-2015, 2015.

Vrekoussis, M., Wittrock, F., Richter, A. and Burrows, J. P.: Temporal and spatial variability of glyoxal as observed from space, Atmos. Chem. Phys., 9(13), 4485–4504 [online] Available from: http://www.atmos-chem-phys.net/9/4485/2009/, 2009.

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?, Atmos. Chem. Phys., 10(21), 10145–10160, doi:10.5194/acp-10-10145-2010, 2010.

Wang, Y., Tao, J., Cheng, L., Yu, C., Wang, Z., Chen, L., Wang, Y., Tao, J., Cheng, L., Yu, C., Wang, Z. and Chen, L.: A Retrieval of Glyoxal from OMI over China: Investigation of the Effects of Tropospheric NO2, Remote Sens., 11(2), 137, doi:10.3390/rs11020137, 2019.

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, Geophys. Res. Lett., 33(16), L16804, doi:10.1029/2006GL026310, 2006.