



ESA Climate Change Initiative

Ozone_cci+

**Product Validation and Intercomparison Report
(PVIR)**

Date: 14.07.2022

Version: 4.0 (Final)

Phase I, Task 4

Deliverable D4.1

WP Manager: J.-C. Lambert

WP Manager Organization: BIRA-IASB

Other partners:

VALT: AUTH, BIRA-IASB

EOST: BIRA-IASB, DLR-IMF, KNMI, RAL, ULB, FMI, IUP-UB



DOCUMENT PROPERTIES

Title Product Validation and Intercomparison Report (PVIR)
Reference Ozone_cci+_PVIR_4.0
Internal references Ozone_cci+_D4.1
Issue / Revision 4.0 / 3
Status Final
Date of issue 14.07.2022
Document type QA/Validation Report

FUNCTION		NAME	DATE
LEAD AUTHOR	VALT Leader	Jean-Christopher Lambert, BIRA-IASB	13/07/2022
CONTRIBUTING AUTHORS	VALT	Dimitris Balis, AUTH Katerina Garane, AUTH José Granville, BIRA-IASB Daan Hubert, BIRA-IASB Arno Keppens, BIRA-IASB Mariliza Koukouli, AUTH Tijl Verhoelst, BIRA-IASB	13/07/2022
	EOST	Christophe Lerot, BIRA-IASB Diego Loyola, DLR-IMF Alexei Rozanov, IUP-UB Richard Siddans, RAL Viktoria Sofieva, FMI Ronald van der A, KNMI Catherine Wespes, ULB	13/07/2022
REVIEWED BY	Science Leader	Michel Van Roozendael, BIRA-IASB	14/07/2022
ISSUED BY	VALT Leader	Jean-Christopher Lambert, BIRA-IASB	14/07/2022
ACCEPTED BY	ESA CCI Officer	Christian Retscher, ESA/ESRIN	14/07/2022



Table of Contents

EXECUTIVE SUMMARY	5
1 Introduction	6
1.1 Purpose and scope	6
1.2 Document overview	6
2 Climate Research Data Package (CRDP).....	7
2.1 Nadir total ozone column data sets	7
2.2 Nadir ozone profile data sets	9
2.3 Limb ozone profile data sets	10
3 ECV Validation Methodology.....	12
3.1 General principles of the validation process	12
3.2 Compliance with user requirements (URD v3.1).....	12
3.2.1 Total ozone data product requirements	14
3.2.2 Nadir ozone profile data product requirements	15
3.2.3 Limb ozone profile data product requirements	17
3.3 Information content and sensitivity	18
3.4 Validation of individual components of ECV processing chain	19
3.5 Confrontation with independent reference measurements.....	19
3.5.1 Total ozone column validation data sources	19
3.5.2 Nadir ozone profile validation data sources.....	24
3.5.3 Limb ozone profile validation data sources.....	25
3.5.4 Error budget of the comparison of atmospheric data.....	26
4 Validation of Total Ozone Data Products	27
4.1 Scope and generalities.....	27
4.2 Level-2 total ozone retrieved with GODFIT v4	27
4.2.1 TROPOMI/S5P Level-2 total ozone	28
4.2.2 GOME-2C Level-2 total ozone	31
4.2.3 Overview of the Level-2 systematic bias and its variations.....	33
4.2.4 Inter-sensor stability.....	33
4.2.5 Seasonality.....	33
4.2.6 Long-term stability.....	36
4.2.7 Dependence on influence quantities.....	39
4.2.8 Summary and compliance with user requirements	41
4.3 Level-3 merged gridded total ozone	45
4.3.1 Level-3 GTO-ECV	45
4.3.2 Level-3 validation results and discussion	47
4.3.3 Summary and compliance with user requirements	49
4.4 Level-2 total ozone from IASI-A, -B and -C	52
4.4.1 Systematic bias and its variations.....	53
4.4.2 Dependency on various parameters of the retrieval algorithm.....	57
4.4.3 Summary and compliance with user requirements	59
4.5 Level-4 assimilated total ozone	61
4.5.1 The MSR Level-4 total ozone product	61
4.5.2 Level-4 validation results and discussion	62
4.5.3 Summary and compliance with user requirements	64
5 Validation of Nadir Ozone Profile Data Products	65
5.1 Nadir ozone profile CRDP	65
5.2 Validation approach	67
5.2.1 Information content studies.....	67
5.2.2 FRM comparisons	68
5.3 Validation results.....	69



5.3.1	Consistency of GOME-2A/B/C instrument retrievals	69
5.3.2	Assessment of the RAL v2 to v3 retrieval updates	72
5.3.3	Validation of the extended OMI time series	75
5.3.4	Consistency of IASI-A/B/C instrument retrievals.....	75
5.3.5	Assessment of FORLI retrieval update to v20191122	76
5.4	Results discussion and conclusions	82
6	Validation of Limb Ozone Profile Data Products	84
6.1	Introduction.....	84
6.2	Harmonized validation methodology.....	84
6.3	Conversion of profile representations.....	85
6.4	Level-2 limb profile products.....	86
6.4.1	Validation method	86
6.4.2	Envisat GOMOS ALGOM2s v1.....	87
6.4.3	Envisat GOMOS Bright Limb v1.2	92
6.4.4	Envisat MIPAS IMK/IAA v7.....	97
6.4.5	Envisat MIPAS IMK/IAA v8.....	102
6.4.6	Envisat SCIAMACHY UBr v3.5	107
6.4.7	Odin OSIRIS v5.10	112
6.4.8	SCISAT-1 ACE-FTS v3.5/v3.6.....	116
6.4.9	SCISAT-1 ACE-FTS v4.1	121
6.4.10	Suomi-NPP OMPS-LP USask-2D v1.0.2	125
6.4.11	Suomi-NPP OMPS-LP USask-2D v1.1.0	130
6.4.12	ERBS SAGE II v7.0.....	135
6.4.13	UARS HALOE v19.....	140
6.4.14	TIMED SABER 9.6 μ m v2.0	145
6.4.15	EOS-Aura MLS v4.2	150
6.4.16	SPOT-4 POAM III v4	155
6.4.17	Meteor-3M SAGE III v4	160
6.4.18	ISS SAGE III v5.1	165
6.4.19	Summary Graphs	170
6.5	Level-3 limb profile products.....	173
6.5.1	Evaluation method	173
6.5.2	Validation results	174
6.5.3	Compliance with user requirements	181
7	Comparison error budget and compliance criteria	182
7.1	Consistency of two measurements	182
7.2	Consistency of a set of N co-located measurement pairs	183
7.3	Application to S5p-TROPOMI total O3	183
7.3.1	A case study: S5p vs. the Dobson at Brisbane, AUS	183
7.3.2	Network-wide results	184
7.4	Caveats and prospects.....	186
8	References	187
8.1	Applicable documents	187
8.2	Reference documents.....	187
8.2.1	User requirements.....	187
8.2.2	International standards and frameworks.....	187
8.2.3	Validation measurements, validation methods and assessments	188
9	Terms and definitions	195
9.1	Terminology.....	195
9.2	Abbreviations and acronyms.....	199



EXECUTIVE SUMMARY

The Ozone_cci+ project is the successor of Ozone_cci as part of ESA's Climate Change Initiative (CCI). The Validation Team (VALT) has developed a Product Validation Plan (PVP) translating user requirements into validation requirements, in order to ensure independent and traceable validation of the Ozone_cci+ data products and verification of compliance with the user requirements. This Product Validation and Intercomparison Report (PVIR) version 4.0 reports on the quality of the Climate Research Data Package (CRDP). For each of the Essential Climate Variable (ECV) data records provided by the project, the PVIR provides users with detailed validation results, with a list of quality indicators enabling the verification of fitness-for-purpose of the data for their own application, and with an assessment of the compliance of the CRDP with user requirements established by the Climate Research Group (CRG) based on their own research needs and on more generic needs formulated by international climate research and monitoring bodies like GCOS.



1 Introduction

1.1 Purpose and scope

The current Ozone_cci+ Climate Research Data Package (CRDP) includes records of total ozone columns, nadir-based ozone profiles, and limb-based ozone profiles. Requirements for the necessary quality assessment of the CRDP datasets are detailed in the Ozone_cci+ Product Validation Plan (PVP, [RD3]) established by a validation team working independently of the algorithm development teams. Based on the PVP, the present Product Validation and Intercomparison Report (PVIR) provides users and product developers of the CRDP with geophysical validation results and with a list of quality indicators enabling the verification of the fitness-for-purpose of the data. In particular, the PVIR discusses the compliance of the individual CRDP datasets with user requirements formulated by GCOS and the project's Climate Research Group (CRG) in a dedicated User Requirements Document (URD, [RD8]). This PVIR will be updated in a future phase of this project as improved and new data products and possibly validation approaches are developed.

1.2 Document overview

The Ozone_cci+ Product Validation and Intercomparison Report is organised as follows:

- Section 2 introduces the CRDP datasets addressed in this report.
- Section 3 describes the ECV validation methodology: Generic principles of the validation process, study of compliance with user requirements, information content and sensitivity studies, and confrontation to independent and traceable reference measurements.
- Section 4 describes validation results and compliance assessment for the total ozone ECV.
- Section 5 describes validation results and compliance assessment for the nadir-based ozone profile ECV.
- Section 6 describes validation results and compliance assessment for the limb-based ozone profile ECV.
- Section 7 discusses the comparison error budget and compliance criteria.
- Section 8 lists applicable and reference documents.
- Section 9 defines the applicable terminology.

For each ECV data product, the results are reported as follows:

- A description of the reference measurements used for independent ECV validation.
- A description of the preparation of satellite and reference measurements, including quality control procedures applied for the selection of the most appropriate data, information on the uncertainties associated to them, co-location criteria applied, data manipulations applied to convert data units and representation systems...
- A description of the match-up analyses performed on the derived ECV products against the selected reference observations.
- A detailed analysis of the uncertainty of the ECV products with reference to the independent validation data.
- Statement of compliance with user requirements formulated in the project's User Requirement Document [RD8].



2 Climate Research Data Package (CRDP)

The Climate Research Data Package (CRDP) generated in the framework of the Ozone_cci+ project contains a list of ozone column and ozone profile data sets. The database can be accessed through the freely accessible ftp site ftp://cci_web@ftp-ae.oma.be/esacci/ozone, and Level-3 products through the CCI Open Data Portal (<http://cci.esa.int/data>) or at the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu>). The data package is organised in three families of ozone data products: Total ozone data products (TC=Total Column), Nadir ozone profile data products (NP=Nadir Profile) and Limb ozone profile data products (LP=Limb Profile). All data sets are delivered in Net-CDF-CF format and are compliant with CCI rules. The Ozone_cci+ data products validated in this document are listed in the tables below. A full description of the retrieval algorithms and retrieval settings is given in the associated Algorithm Theoretical Basis Document (ATBD) [RD4].

2.1 Nadir total ozone column data sets

Table 2.1 - Ozone_cci+ CRDP total ozone column data products.

Product level	Product ID	Sensor	Product description	Provider	Time coverage	Validation report
Level 2	TC_L2_GOME	ERS-2 GOME	Harmonized GODFIT multi-sensor prototype level 2 data	BIRA-IASB	Lifetime (1995-2011); global coverage lost after June 2003	Section 4.2
	TC_L2_SCIA	Envisat SCIAMACHY		BIRA-IASB	Lifetime (2002-2012)	Section 4.2
	TC_L2_GOME2A	MetOp-A GOME-2		BIRA-IASB	Lifetime (since 2007)	Section 4.2
	TC_L2_GOME2B	MetOp-B GOME-2		BIRA-IASB	Lifetime (since 2013)	Section 4.2
	TC_L2_OMI	Aura OMI		BIRA-IASB	Lifetime (since 2004)	Section 4.2
	TC_L2_OMPS	SNPP OMPS		BIRA-IASB	Lifetime (since 2004)	Section 4.2
	TC_L2_TROPOMI	S5P TROPOMI		BIRA-IASB	Lifetime (since 2018)	Section 4.2
	TC_L2_GOME2C	MetOp-C GOME-2		BIRA-IASB	Lifetime (since 2019)	Section 4.2
Level 3	TC_L3_MRG a.k.a. GTO-ECV	Combined	GOME, SCIAMACHY, GOME-2A/B and OMI merged prototype level 3 harmonized data	DLR	1995-2019	Section 4.3
Level 2		IASI MetOp-A	FORLI-O3 v20191122	ULB	05/2015 – 12/2020	Section 4.4
Level 2		IASI MetOp-B	FORLI-O3 v20191122	ULB	05/2015 – 12/2020	Section 4.4
Level 2		IASI MetOp-C	FORLI-O3 v20191122	ULB	09/2019 – 12/2020	Section 4.4



Product level	Product ID	Sensor	Product description	Provider	Time coverage	Validation report
Level 4	MSR	Combined	TM3-DAM v3.3 assimilated product, with data from GOME, SCIAMACHY, OMI, GOME-2A/B, BUV-Nimbus4, TOMS-Nimbus7, TOMS-EP and SBUV-7, -9, -11, -14, -16, -17, -18, -19	KNMI	1970-2020	Section 4.5



2.2 Nadir ozone profile data sets

Table 2.2 - Ozone_cci+ CRDP nadir ozone profile data products.

Product level	Product ID	Sensor	Product description	Provider	Time coverage	Validation report
Level 2	NP_L2_GOME	ERS-2 GOME	CCI algorithm version RAL 3.01 with profiles on fixed pressure levels from SPARC DI	RAL	1996-2011	Section 5.3
	NP_L2_SCIA	Envisat SCIAMACHY	CCI algorithm RAL 3.00 with profiles on fixed pressure levels from SPARC DI	RAL	2002-2012	Section 5.3
	NP_L2_GOME2A	Metop-A GOME-2	CCI algorithm RAL 3.00 and 3.02 with profiles on fixed pressure levels from SPARC DI	RAL	2007-2021	Section 5.3
	NP_L2_GOME2B	Metop-B GOME-2	CCI algorithm RAL 3.03 and 3.05 with profiles on fixed pressure levels from SPARC DI	RAL	2014-2021	Section 5.3
	NP_L2_GOME2C	Metop-C GOME-2	CCI algorithm RAL 3.00 with profiles on fixed pressure levels from SPARC DI	RAL	2020-2021	Section 5.3
	NP_L2_OMI	Aura OMI	CCI algorithm RAL 2.14 with profiles on fixed pressure levels from SPARC DI	RAL	2004-2021	Section 5.3
	NP_L2_TROPOMI	Sentinel-5p TROPOMI	TBD	RAL	TBD	/
	NP_L2_GOME2-IASI	Metop-A/B/C GOME-2 and IASI	TBD	RAL	TBD	/
	NP_L2_IASIA	Metop-A IASI	FORLI 20151001 and 20191122 algorithm on fixed altitude levels	ULB-LATMOS	2008-2021	Section 5.3
	NP_L2_IASIB	Metop-B IASI	FORLI 20151001 and 20191122 algorithm on fixed altitude levels	ULB-LATMOS	2013-2021	Section 5.3
	NP_L2_IASIC	Metop-C IASI	FORLI 20151001 and 20191122 algorithm on fixed altitude levels	ULB-LATMOS	2019-2021	Section 5.3
Level-3	NP_GOP-ECV	TBD	TBD	DLR	1996-2020	/



2.3 Limb ozone profile data sets

Table 2.3 - Ozone_cci+ CRDP limb ozone profile data products.

Product level	Product ID	Sensor	Product description	Provider	Time coverage	Validation report
HARMonized dataset of OZone profiles (HARMOZ) Level 2	LP_L2_GOMOS	Envisat GOMOS	Individual profiles of ozone mole concentrations on a common pressure or geometric altitude grid, and auxiliary information to convert to volume mixing ratio and/or geometric altitude/pressure	FMI	Entire mission; data screened for outliers (filtered data)	Section 6.4.2
	LP_L2_GBL	Envisat GOMOS		FMI		Section 6.4.3
	LP_L2_MIPAS	Envisat MIPAS		IMK/IAA		Sections 6.4.4 6.4.5
	LP_L2_SCIA	Envisat SCIAMACHY		U Bremen		Section 6.4.6
	LP_L2_OSIRIS	Odin OSIRIS		U Saskatchewan		Section 6.4.7
	LP_L2_ACE	SciSat ACE FTS		U Toronto		Sections 6.4.8 6.4.9
	LP_L2_OMPS	Suomi-NPP		U Saskatchewan		Sections 6.4.10 6.4.11
	LP_L2_SAGE-II	ERBS SAGE II		NASA-LaRC, U Bremen		Section 6.4.12
	LP_L2_HALOE	UARS HALOE		NASA-LaRC, U Bremen		Section 6.4.13
	LP_L2_SABER	TIMED SABER		NASA-LaRC, U Bremen		Section 6.4.14
	LP_L2_MLS	Aura MLS		NASA-JPL, U Bremen		Section 6.4.15
	LP_L2_POAM-III	SPOT-4 POAM III		NASA, U Bremen		Section 6.4.16
	LP_L2_SAGE-III/M3M	Meteor-3M SAGE III		NASA-LaRC, U Bremen		Section 6.4.17
	LP_L2_SAGE-III/ISS	ISS SAGE III		NASA-LaRC, U Bremen		Section 6.4.18
Level 3	LP_L3_GOMOS	Envisat GOMOS	Monthly zonal means for each individual instrument, on native vertical grid of instrument	FMI	Entire mission	Section 6.5.2
	LP_L3_MIPAS	Envisat MIPAS				
	LP_L3_SCIA	Envisat SCIAMACHY				
	LP_L3_OSIRIS	Odin OSIRIS				
	LP_L3_ACE	SciSat ACE FTS				
	LP_L3_OMPS	SNPP OMPS-LP				
	LP_L3_SAGE-II	ERBS SAGE II				
	LP_L3_HALOE	UARS HALOE				
	LP_L3_SABER	TIMED SABER				
LP_L3_MLS	Aura MLS					



Product level	Product ID	Sensor	Product description	Provider	Time coverage	Validation report
	LP_L3_MERGED_SAGE_CCI_OMPS	SAGE II, OSIRIS, GOMOS, MIPAS, SCIAMACHY, ACE-FTS, OMPS-LP	Monthly mean anomaly in 10° latitude zones		1984-2020	
	LP_L3_MEGRIDOP	OSIRIS, GOMOS, MIPAS, SCIAMACHY, Aura MLS, OMPS-LP	Monthly mean anomaly in 10° latitude x 20° longitude cells		2001-2020	



3 ECV Validation Methodology

3.1 General principles of the validation process

The Ozone_cci+ Product Validation Plan [RD3] describes the validation protocol applied in this assessment. The prime objective of the Ozone_cci+ project is the production of ECV data products responding to the needs of the climate research community, represented by the Ozone_cci+ Climate research Group (CRG) and the CCI Climate Modelling User Group (CMUG). Every ECV data set produced by the project needs to be validated against the official user requirements formulated in the Ozone_cci+ User Requirement Document (URD, [RD8]). In the following section, we summarize the user requirements applicable to the present validation study. The translation of these user requirements into validation requirements is described in the Ozone_cci+ Product Validation Plan (PVP, [RD3]). The geophysical validation of ECV data products delivered in the CRDP relies primarily on comparisons with ground-based reference measurements. These comparisons are reported in Sections 4 to 6. The reference measurements used in this study are summarised in Section 3.5. In preparation of the comparisons, the data sets must undergo a suite of data manipulations, including data filtering based on, e.g., quality flags, harmonisation of coordinate systems and of units, reduction of differences in vertical and horizontal smoothing, selection of co-locations meeting appropriate criteria... These operations depend on the ECV data product and associated retrieval algorithms; therefore they are described in the respective sections reporting the comparisons. Prior to the data comparisons, the characterisation of the information content of the data products and their sensitivity to the real atmosphere may be required. This is definitely the case for a proper interpretation of nadir ozone profile data, for which the final data product is a mix of real contributions from the measurement and of a priori constraints. This aspect is addressed in Section 3.3.

3.2 Compliance with user requirements (URD v3.1)

The Ozone_cci+ User Requirement Document (URD) [RD8] defines climate user requirements based on the ozone requirements of the Global Climate Observing System (GCOS), the CCI Climate Modelling User Group (CMUG) [RD5], the Integrated Global Atmospheric Chemistry Observation theme (IGACO) of the Integrated Global Observing Strategy (IGOS) [RD7], and the WMO observational requirements [RD9]. They are summarised hereafter. These URD requirements were translated into validation requirements in the Product Validation Plan (PVP) [RD3] established by the Ozone_cci+ Validation Team (VALT).

The first category of user requirements addresses classical error bars. In the case of total ozone column TOC (expressed in DU) the error will be given as a delta total ozone value in DU (δTOC), usually equal to a few percent, such that $\text{TOC} \pm \delta\text{TOC}$ represents a symmetric 68 % confidence interval. This δTOC value contains a systematic term and a random term, corresponding to classical bias and precision (1σ standard deviation or equivalent) estimates. Validation is expected to verify the accuracy of ex-ante estimates of the systematic bias and precision provided by the ECV retrieval teams. This verification must further ensure that these quality indicators, which usually vary with several parameters of the measurement and the retrieval, remain within the acceptable ranges defined in URD.

In the case of ozone profiles two error bars are required, one representing an altitude range (requirement of $\pm 500\text{m}$ for limb profile retrievals), the other representing a volume mixing ratio range (requirements between 8% and 20%), and both representing a symmetric 68 % confidence interval. Assessment of the error bar on altitude for nadir ozone profile data requires analysis of information content (e.g., calculation of centroids and Backus-Gilbert spread of the vertical averaging kernels, [RD25]). Details will be addressed in dedicated sections.



The second category of user requirements addresses (i) the temporal and spatial domains over which, and (ii) the associated temporal and spatial resolutions at which, data quality must meet the first category of user requirements:

- Temporal domain and sampling: continuous coverage with 3 days of observation frequency over the decadal range and beyond, with maximum uncertainty on interannual variability, annual cycle and shorter term variability ranging from 2-3% for total ozone data up to 20% for tropospheric ozone data.
- Temporal stability: Long-term stability of 1%-3%/decade to allow trend detection.
- Geographical domain: global, regional, latitude-height monthly mean cross-sections.
- Horizontal resolution requirements: from 20 km to 300 km depending on the ECV.
- Vertical range and sensitivity: requirements reflect the vertical structure of ozone changes, namely total ozone column (TOC), and ozone in the lower troposphere (LT), upper troposphere (UT), lower stratosphere (LS), upper stratosphere and mesosphere (USM).
- Vertical resolution: depending on the ECV.

Other user requirements fall rather into categories of product specifications:

- Level of the ECV data set: off-line homogenized Level-2 time series for process evaluations on time scales spanning from hours/days to months/years, and homogenized multi-instrument long-term data sets for ozone-climate interactions (Level-3 and Level-4).
- Continuity of user requirements between data levels, e.g., aggregated multi-sensor Level-3 products should retain Level-2 requirements as much as possible. At least, Level-3 products should not be homogenized/degraded to the instrument with the lowest accuracy over the targeted time period.
- Requirements for ancillary data: cloud information per pixel (including cloud fraction, cloud height, cloud albedo) and surface information per pixel (surface albedo).
- Data format and metadata requirements.
- Visualisation requirements.

Compliance with requirements on observation frequency and geographical domain is straightforward to verify through visualisation of the data sets, a study hereafter referred to as dataset content study. Compliance with requirements on spatial resolution and spatial sampling need visualisation of the data and analysis of information content (e.g., calculation of centroids and Backus-Gilbert spread and use of cross-correlation techniques). Compliance with more specific requirements, especially requirements peculiar to Level-3/4 data products, e.g. in terms of actual geographical coverage and of point-to-zone representativeness, may need the use of statistical methods based on global model results. In addition to validation studies and quality checks performed by the validation teams (VALT) and by the ECV producers (EOSTs), user feedbacks provide valuable input for the assessment of effective usability of the data product. The latter quality checks are reported in another document, the Climate Assessment Report (CAR) [RD70].

Hereafter we reproduce the user requirements as described in Tables 5 to 10 of version 3.1 of the URD, against which Ozone_cci+ ECV data products have to be verified and/or validated. For each ECV the tables in this document display specific requirements on the data, its characteristics and its errors (Table 3.1, Table 3.3 and Table 3.5), and requirements on the data format and associated metadata (Table 3.2, Table 3.4 and Table 3.6).



3.2.1 Total ozone data product requirements

Table 3.1 - Product requirements for total ozone column data. Achievable and future target requirements are given, separated by a ‘–’ (adapted from URD v3.1, 2021).

Quantity	Driving Research topic	Geographical Zone		
		Tropics	Mid-latitudes	Polar region
Global horizontal resolution	Evolution of the ozone layer (radiative forcing); Seasonal cycle and interannual variability; Short-term variability*	20 – 100 km	20 – 50/100 km	20 – 50/100 km
Observation frequency	Evolution of the ozone layer (radiative forcing); Seasonal cycle and interannual variability; short-term variability*	Daily – weekly	Daily – weekly	Daily – weekly
Time period	Evolution of the ozone layer (radiative forcing)	(1980-2010) 1995-2011	(1980-2010) 1995-2011	(1980-2010) 1995-2011
Uncertainty	Evolution of the ozone layer (radiative forcing)	2% (7 DU)	2% (7 DU)	2% (7 DU)
Uncertainty	Seasonal cycle and interannual variability; Short-term variability*	3% (10 DU)	3% (10 DU)	3% (10 DU)
Stability (after corrections)	Evolution of the ozone layer (1980-2010 trend detection; radiative forcing)	1 – 3 % / decade (1995-2011)	1 – 3 % / decade (1995-2011)	1 – 3 % / decade (1995-2011)

* Short-term variability includes: Exchange of air masses, streamers, regime studies.

Table 3.2 - Data format and metadata requirements for total ozone (adapted from URD v3.1, 2021)

Data feature	Requirement
Data format	Net-CDF [RD20]
Data conventions	CF
Data units	Total column (in DU; number of molecules per area or equivalent)
Error	Total area
Error characteristics (optional)	Total uncertainty and its subdivision per pixel into: - contribution measurement noise; - contribution of a priori uncertainties; - contribution of estimated spectroscopic uncertainty
Averaging kernels	Yes for Level-2
Full covariance matrix included ?	No
A priori data	Yes, per pixel
Quality flag	1: high quality data 2: contaminated data 3: missing value
Visualisation	Basic browsable archive visualisation (daily global maps; local/latitudinal time series of monthly means)



3.2.2 Nadir ozone profile data product requirements

Table 3.3 - Product requirements for nadir-based ozone profile Climate Data Records (CDRs). The ozone profile requirements are for ozone products in terms of (partial-column mean) mixing ratios. The tropospheric altitude domain extends from the surface to the tropopause defined by an ozone concentration of 150 ppbv; the UT/LS extends from about 5 to 30 km, and the middle atmosphere extends from about 30 to 60 km altitude. The required coverage is global. Achievable and future target requirements are given, separated by a ‘-’. The first number is the future target. Note: requirements have been updated in Ozone_cci+. (adapted from URD v3.1, 2021)

Quantity	Driving Research topic	Height range		
		Troposphere	UT/LS	Middle Atmosphere
Horizontal resolution	Regional differences in evolution of the ozone layer and tropospheric ozone burden (radiative forcing); Seasonal cycle and interannual variability; Short-term variability*	20 – 200 km	20 – 200 km	200 – 400 km
Vertical resolution	Height dependence of evolution of the ozone layer and the tropospheric ozone burden (radiative forcing); Seasonal cycle and interannual variability; Short-term variability*	6 km – tropospheric column	6 km – partial column	6 km – partial column
Observation frequency	Evolution of the ozone layer and the tropospheric ozone burden (radiative forcing); Seasonal cycle and interannual variability; Short-term variability*	Daily – weekly	Daily – weekly	Daily – weekly
Time period	Evolution of the ozone layer and tropospheric ozone burden (radiative forcing)	(1980-2010) – (1996-2010)	(1980-2010) – (1996-2010)	(1980-2010) – (1996-2010)
Accuracy	Evolution of the ozone layer and tropospheric ozone burden (radiative forcing)	8%	8%	8%
Accuracy	Seasonal cycle and interannual variability; Short-term variability*	16%	16% (< 20 km) 8% (> 20 km)	8%
Stability	Evolution of the ozone layer and tropospheric ozone burden (radiative forcing); trends	1 – 3% / decade	1 – 3% / decade	1 – 3% / decade

* Short-term variability includes: Exchange of air masses, streamers, regime studies.



Table 3.4 - Data requirements for nadir-based ozone profile Climate Data Records (adapted from URD v3.1, 2021)

Data feature	Requirement
Data format	Net-CDF
Data conventions	CF
Data units	Ozone mixing ratio (optional: also in partial ozone column and/or with co-located temperature profile)
Error characteristics	Total accuracy and its subdivision per pixel and per layer into: - contribution measurement noise; - contribution smoothing error - contribution of A Priori uncertainties;
Number of layers	To be chosen for optimal accuracy (not too few for information content, not too many by degrading the accuracy per layer)
Averaging kernels included ?	Yes, per pixel
Full covariance matrix included ?	Yes, per pixel
A priori data included ?	Yes, per pixel
Flags	Quality per pixel (good, bad, uncertain); Pixel type; Snow/ice; Sun glint; Solar Eclipse; South-Atlantic Anomaly
Visualisations	Basic browsable archive visualisation (profile cross section per orbit; monthly maps at standard pressure levels; local/latitudinal time series of monthly means at standard pressure levels)



3.2.3 Limb ozone profile data product requirements

Table 3.5 - Product requirements for limb-based ozone profile Climate Data Records (CDRs). The ozone profile requirements are for ozone products in terms of (partial-column mean) mixing ratios. The lower stratosphere (LS) extends from the tropopause (defined as ozone > 150 ppbv) to about 30 km, and the middle atmosphere extends from about 30 to 60 km altitude. The required coverage is global. Achievable and future target requirements are given, separated by a ‘-’. The first number is the future target. (adapted from URD v3.1, 2021)

Quantity	Driving Research topic	Height Range	
		Lower Stratosphere	Middle Atmosphere
Horizontal resolution	Regional differences in the evolution of the ozone layer (radiative forcing); Seasonal cycle and interannual variability; Short-term variability	100 – 200 km	200 – 400 km
Vertical resolution	Height dependence of evolution of the ozone layer (radiative forcing); Seasonal cycle and interannual variability; Short-term variability	1 – 2 km	2 – 4 km
Observation frequency	Seasonal cycle and interannual variability; short-term variability	Daily – weekly	Daily – weekly
Time period	Evolution of the ozone layer (radiative forcing)	(1980-2010) – (2003-2010)	(1980-2010) – (2003-2010)
Uncertainty in height attribution	Evolution of the ozone layer (radiative forcing), Seasonal cycle and interannual variability; Short-term variability	±500 m	±500 m
Uncertainty on mixing ratio	Evolution of the ozone layer (radiative forcing)	8%	8%
Uncertainty on mixing ratio	Seasonal cycle and interannual variability; Short-term variability	16 % (<20 km) 8% (>20 km)	8 %
Stability	Evolution of the ozone layer (radiative forcing); trends	1 – 3 % / decade	1 – 3 % / decade



Table 3.6 - Data format and metadata requirements for limb-based ozone profile requirements (adapted from URD v3.1)

Data feature	Requirement
Data format	Net-CDF [RD20]
Data conventions	CF
Data units	Ozone mixing ratio (optional: also in partial ozone column and/or with co-located temperature profile)
Error characteristics	Total accuracy and its subdivision per profile per layer into: - contribution measurement noise; - contribution horizontal smoothing error - contribution pointing accuracy - contribution of A Priori uncertainties;
Averaging kernels included ?	Yes, per profile
Full covariance matrix included ?	Yes, per profile
A priori data included ?	Yes, per profile
Flags	Quality per profile per layer (good, bad, uncertain); Cloud contamination; Solar Eclipse; South-Atlantic anomaly
Visualisation	Basic browsable archive visualisation (profile cross section per orbit; monthly maps at standard pressure levels; local/latitudinal time series of monthly means at standard pressure levels)

3.3 Information content and sensitivity

A key aspect in the validation of usability (the verification of “fitness for purpose” of a data product) is the characterisation of the information content of the data product. The retrieval of geophysical quantities from remote sounding measurements usually uses a set of a priori constraints, e.g., in the form of an assumed range of atmospheric profile shape around a first guess. Such constraints mix somehow in the retrieved quantities with the information really contributed by the measurement [RD73]. When a climatology is used in the retrieval, e.g., at altitudes where the measurement is not or less sensitive due to optically thick clouds or due to too low signal-to-noise ratios, it is important to understand what, in the final product, comes from the climatology and what comes really from the measurement. This kind of validation of the information content can rely on a combination of (1) comparisons with independent reference data sets, especially during events not considered in the climatology, (2) the study of deviations of the retrieved product from the a priori constraints, and (3) sensitivity analysis of the retrieval, e.g. based on a study of the associated averaging kernels and their eigenvectors [RD73, RD59, RD46]. E.g., plotting as a function of altitude the sum of the rows of the averaging kernel matrix associated with a retrieval shows at which altitudes the measurement offers sensitivity to atmospheric concentrations. Similarly, the real information content of the reference measurement itself should be known prior to performing a comparison. Information content studies might be an important aspect of the validation of model runs that have been initialised by climatology or by the output of another model, or that are constrained by a priori boundary conditions. They can also be of relevance in the assessment of data assimilation results when observations outside of a predetermined range are rejected as outliers by the data ingestion scheme, producing in the system a zero information zone similar to the dead band or neutral zone used in voltage regulators and controllers to avoid unwanted oscillations and disruptions. Information content studies of the Level-2 data are also essential in understanding higher level data products generated by data merging and ensemble approaches.



3.4 Validation of individual components of ECV processing chain

ECV line components are the individual processing blocks by which ECV data products are generated in their interim or final version. For complex processing chains international standards require to validate or at least verify the good performance of every component and the accuracy of its output. Limiting validation to the final data product only is not sufficient. The validation of intermediate data products is highly desirable to avoid, e.g., that the apparently good behaviour of the final data product at the end of the chain hides large compensating errors affecting separate components of the data retrieval. Testing is one of many verification activities intended to confirm that software development output meets its input requirements. Other verification activities include various static and dynamic analyses, code and document inspections, walkthroughs, and other techniques. Most of these verification activities have been performed by the EOSTs developing and producing the data products, and are reported in the associated Algorithm Theoretical Basis Documents (ATBDs) [RD4].

3.5 Confrontation with independent reference measurements

The performance of calibration procedures, retrieval algorithms and data merging systems, and the quality of the resulting ECV products is primarily assessed by comparison with traceable reference measurements supposed to provide the “true” atmospheric state. A key aspect is the appropriate selection of the reference data sets. The quality, traceability and fitness-for-purpose of the latter are essential to allow proper, unbiased and independent validation. Reference measurements must be well documented, and procedures must exist to ensure their quality control on the long term, as it is the case, e.g., within international ground-based networks where data acquisition and QA/QC are regulated by protocols.

Ground-based reference measurements of the total column and vertical distribution of atmospheric ozone are performed by networks of instruments contributing to WMO’s Global Atmosphere Watch (GAW) [RD22]. In the Ozone_cci+ project, ground-based data sets suitable for the validation of ECV products are collected from complementary instruments archiving routinely their data to the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) and the Data Host Facility (DHF) of the Network for the Detection of Atmospheric Composition Change (NDACC). Details including data acquisition protocols, data quality estimates and data access conditions, are available on the web portals of the data archives (<http://woudc.org> and <http://ndacc.org>, respectively) and summarised in the Data Access Requirement Document (DARD) [RD6]. Additionally, satellite data sets of documented quality are also used to extend ground-based validation results to a more global coverage and identify features that cannot be detected by a network like, e.g., geographical patterns.

3.5.1 Total ozone column validation data sources

As described in DARD [RD6], the following measurement data sets are used hereafter as reference for validation studies and/or for cross-comparison studies of the total ozone column data products:

- Ground-based total ozone column (TOC) measurements by Dobson and Brewer ultraviolet spectrophotometers.
- Ground-based total ozone column measurements by UV-visible DOAS spectrometers.

3.5.1.1 Brewer and Dobson measurements

The ground-based measurements database used for this validation report consists of archived Brewer and Dobson total ozone data that are downloaded from the World Ozone and Ultraviolet Radiation Data Centre (<http://www.woudc.org>). WOUDC is one of the World Data Centers which are part of the Global Atmosphere Watch (GAW) program of the World Meteorological Organization (WMO). These data are quality controlled, first by each station before submission and secondly by WOUDC. Brewer and Dobson ultraviolet spectrophotometers rely on the method of differential absorption in the Huggins band where ozone exhibits strong absorption features in the ultraviolet part of the solar spectrum. This technique has been described in detail by the main reference papers [RD49] and references therein.



The Dobson spectrophotometer measures TOC values with a total uncertainty of 2%–3% for solar zenith angles smaller than 75°. Since the International Geophysical Year in 1957, Dobson instruments have been deployed in a worldwide network. Figure 3.1 displays the geographical distribution of Dobson stations used in this study. It is known that Dobson measurements suffer from a temperature dependence of the ozone absorption coefficients used in the retrievals which might account for a seasonal variation in the error of $\pm 0.9\%$ in the middle latitudes and $\pm 1.7\%$ in the Arctic, and for systematic errors of up to 4% (Bernhard et al., 2005, [RD27]). For the purposes of this work, the Dobson ground-based measurements that will be used for the total ozone validation are post-corrected following the methodology described below.

The Brewer grating spectrophotometer is in principle similar to the Dobson. However, it has an improved optical design and is fully automated. The ozone column abundance is determined from a combination of five wavelengths between 306 nm and 320 nm. Since the 1980s, Brewer instruments are part of the ground-based network as well. Figure 3.2 displays the geographical distribution of Brewer stations used in this study. Most Brewers are single monochromators, but a small number of systems are double monochromators with improved stray light performance. Most of the Brewer instruments providing data to the WOUDC repository are operated at Northern Hemisphere stations. There are a few instruments of this type in the Southern Hemisphere, but they are not considered in this study because of their limited spatial representativeness. The uncertainty on total ozone Direct Sun (DS) measurements by a well-maintained Brewer instrument is about 1% (e.g., [RD49]). When Brewer spectrophotometers are regularly calibrated and maintained, the DS TOC records can potentially maintain a precision of 1% over long-time intervals [RD94].

Despite similar performance, small differences within $\pm 0.6\%$ on an average are introduced between the Brewer and Dobson data because of the use of different wavelengths and different temperature dependence for the ozone absorption coefficients [RD82]. The seasonal cycle in atmospheric temperature results in a seasonal variation of the Dobson ozone data, where the contribution of the systematic offset is less than 1% [RD89]. Dobson and Brewer instruments might also suffer from long-term drift associated with calibration changes. Additional problems arise at solar elevations lower than 15°, for which diffuse and direct radiation contributions can be of the same order of magnitude. Therefore, we limit the use of measurements by Dobson and Brewer ultraviolet spectrophotometers, to data acquired up to 80° SZA for Brewers MK-III and MK-IV (double monochromators), and up to 70-75° of SZA for Dobsons and other Brewers (single monochromators).

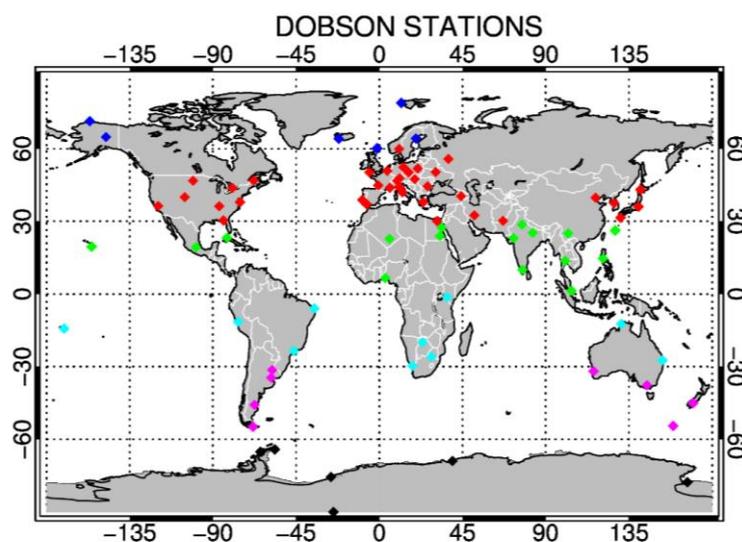


Figure 3.1 - Geographical distribution of Dobson network stations used in this study, colour-coded per latitude band; 90-60°S in black, 60-30°S in purple, 30°S – 0 in cyan, 0-30° N in green, 30-60° N in red and 60-90° N in blue.

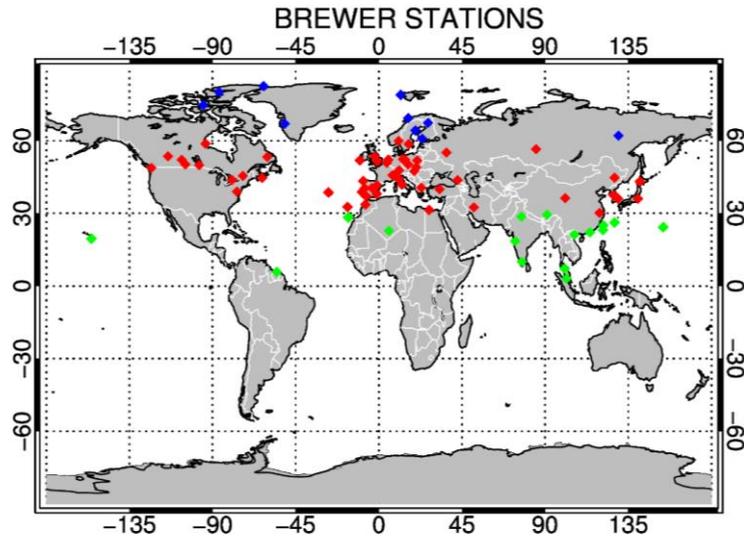


Figure 3.2 - Geographical distribution of Brewer network stations used in this study, colour-coded per latitude band; 0-30° in green, 30-60° in red and 60-90° in blue.

Dobson total ozone column correction for the effective temperature dependence

To account for the Dobson total ozone columns dependence on effective temperature, the methodology introduced by Komhyr et al. (1993) [RD51] was applied to all Dobson datasets, using the Effective Temperature that is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF, <https://www.ecmwf.int>). The station overpass files are available by the Tropospheric Emission Monitoring Internet Service (TEMIS, <https://www.temis.nl/climate/efftemp/overpass.php>).

The post-processing methodology, thoroughly described by Koukouli et al. (2016) [RD53], uses the following formula to calculate the new total ozone values (in DU):

$$O_{3\ new} = O_{3\ standard} [1 - 0.0013(T_{eff} - 226.7)] \quad (3.1)$$

where:

- $O_{3\ standard}$ is the retrieved total ozone column corresponding to the Dobson reference effective temperature (-46.3 °C)
- 226.7 is the Dobson reference effective temperature expressed in Kelvin and
- T_{eff} is the effective temperature from ECMWF

Figure 3.3 shows the effect of the correction on the Hohenpeissenberg station dataset, as an example. In both panels the percentage difference of the corrected TOCs with respect to the initial TOCs is shown. In the upper panel, the seasonal dependence of the differences is depicted, while in the bottom panel the temporal evolution of the differences are shown. As it results from both figures, the effective temperature correction results to higher TOCs by about 1% during winter months and lower TOCs by ~0.8% during summer months with respect to non-corrected measurements, leading to a peak-to-peak difference between the two datasets of ~1.8%. The effect of the correction on the validation results of the OMI TOC data, retrieved by the GODFIT v4 algorithm, is shown in Figure 3.4. The Hohenpeissenberg Dobson spectrophotometer and the OMI overpass data are compared and shown in the form of the time-series of their percentage differences. The right panel shows their comparisons when the non-corrected Dobson data are used as reference, while the left panel uses the effective temperature corrected ground-based measurements as the ground-truth. The statistics, in terms of mean relative bias and seasonality are summarized in Table 3.7. The decrease in the seasonal dependence of the differences is about 2%. Therefore, the effective temperature corrected Dobson TOC measurements (noted as “Dobson_c” in the figures) will be used in this validation study, along with the Brewer TOC observations, as a reference for the validation of Level 2 satellite total ozone products.

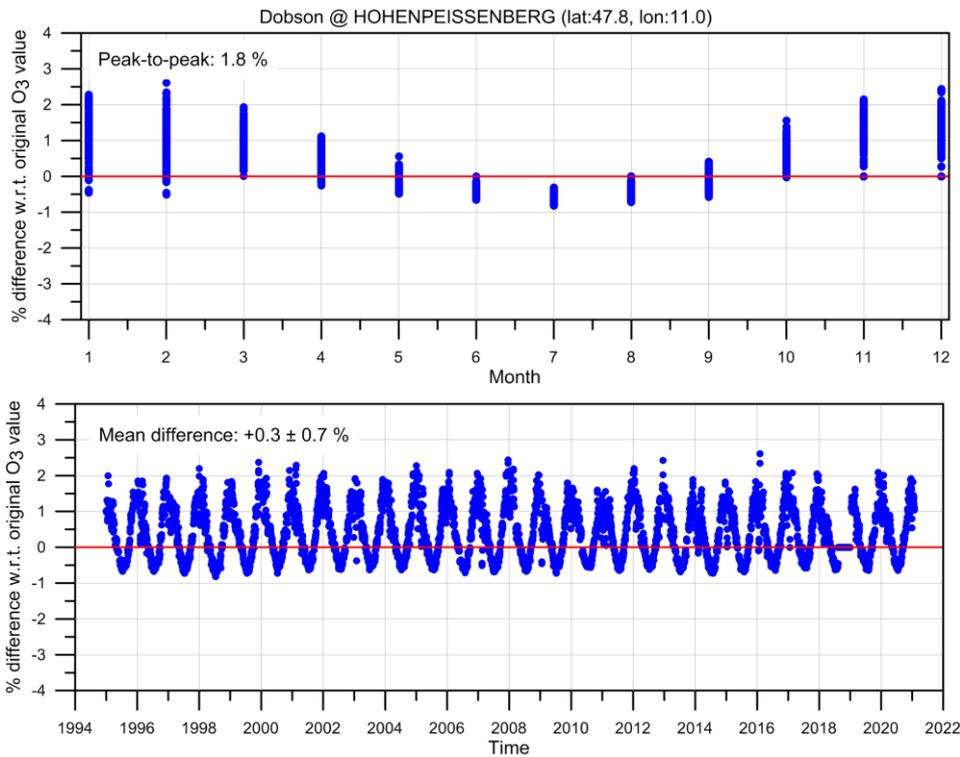


Figure 3.3 - The time-series (bottom panel) and the seasonal dependency (upper panel) of the percentage differences between the effective temperature corrected TOCs with respect to the initial TOCs retrieved by the Hohenpeissenberg Dobson spectrophotometer.

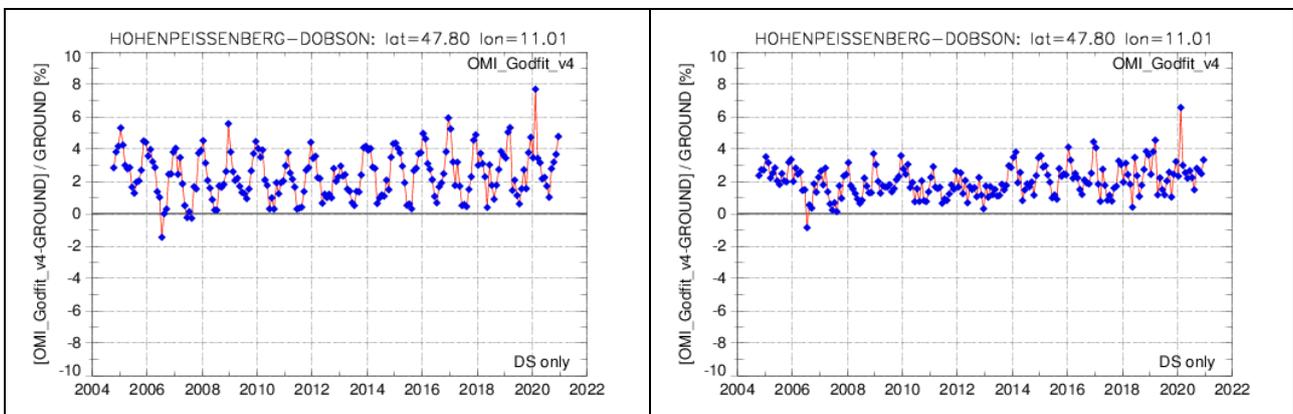


Figure 3.4 - The percentage differences time series of the co-located Hohenpeissenberg Dobson measurements and the OMI overpass data. Right panel: the non-corrected Dobson data are used as reference; left panel: the effective temperature corrected ground-based measurements are used.

Table 3.7 - The statistics of the Hohenpeissenberg Dobson measurements comparison to the OMI TOCs, before and after the effective temperature correction of the ground-based observations

	No temperature correction	Temperature corrected
Mean bias \pm std. dev.:	$+2.2 \pm 1.6$ %	$+1.8 \pm 1.2$ %
Seasonality (peak-to-peak):	~ 4.5 %	~ 2.5 %



3.5.1.2 UV-visible DOAS measurements

Based on the differential optical absorption spectroscopy technique (DOAS), SAOZ is a zenith sky UV-Visible spectrometer measuring O₃ and NO₂ total columns at twilight. This instrument was developed by Pommereau and Goutail [RD72] after discovery of the ozone hole, to enable measurements of O₃ and NO₂ year-round in polar areas even during wintertime, when too low Sun elevations prevent direct Sun measurements of the Brewer and Dobson types. Since then, about 35 UV-visible DOAS instruments have been deployed at all latitudes within the Network for the Detection of Atmospheric Composition Change (NDACC). Figure 3.5 displays the geographical distribution of the UV-visible instruments used in this study. The measurements are acquired at sunrise and sunset between 86°-91° Solar Zenith Angle (SZA), when sensitivity to weak stratospheric absorbers is the highest. The total ozone column is retrieved in the visible Chappuis band. Several characteristics of the zenith-sky twilight DOAS technique add to are the Brewer and Dobson techniques by ensuring year-round measurement capabilities at high latitudes and by being affected by different source of errors: i) the self-calibration provided by the differential absorption cross-sections used in the DOAS analyses, thus offering self-stability properties, identical for all instruments, ii) the negligible temperature dependence of the ozone differential absorption cross-sections in the visible Chappuis band [RD32], iii) the same 86°-91° SZA range used at all latitudes and seasons, thus the independency on SZA, and iv) the weak perturbation of the measurement by clouds below the mean scattering layer of sunlight at twilight (about 10-14 km), except when multiple scattering and consequently light path enhancements occur within thick thunderstorms, dense haze and compact rain/snow showers, whose data are rejected after detection of water vapour and O₄ oxygen dimer enhancements. The current retrieval algorithm version 2 applies the recommendations of the NDACC UV-VIS Working Group [RD42]. With these settings, the largest source of error in the retrieval comes from the Air Mass Factor (AMF) used for converting ozone slant columns into vertical columns, calculated with a radiative transfer code run on the zonal mean profile climatology developed for the TOMSv8 algorithm. Overall, the random uncertainty on the individual total ozone AMF is estimated to 4.7%, which, combined with systematic errors, results in total uncertainty on ozone retrievals to 5.9% [RD42]. When Dobson data are corrected for their temperature dependence, the average difference with daily SAOZ total ozone columns is within -0.8 ± 3.8 % with the Dobson at OHP (Southern France) and 1.9 ± 3.2 % with the Brewer in Sodankylä (Finland).

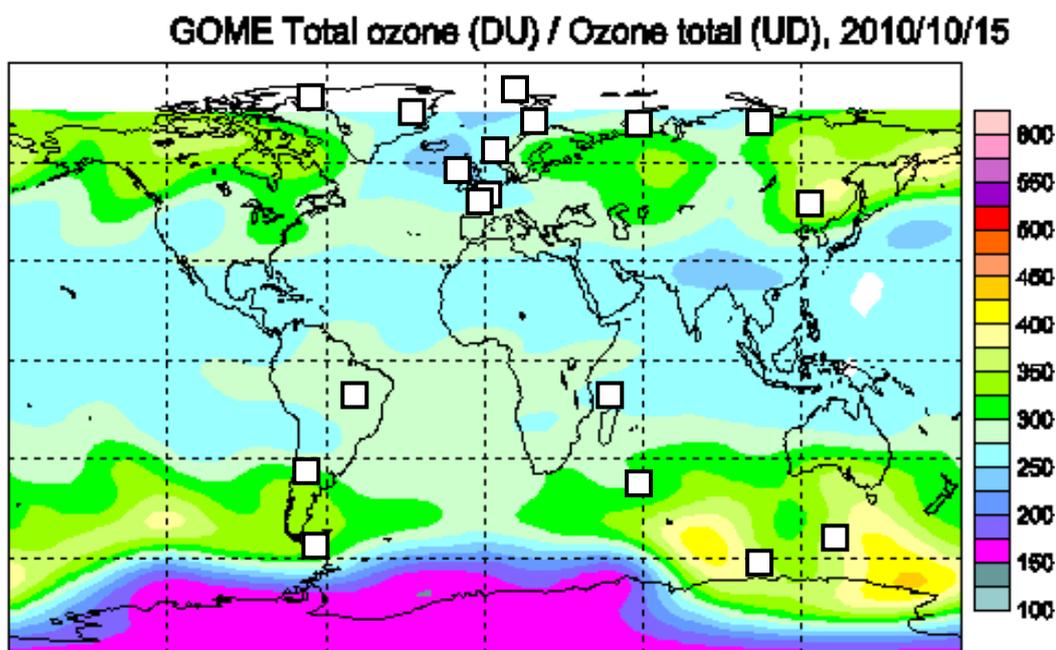


Figure 3.5 - Geographical distribution of NDACC UV-visible DOAS spectrometers used in this study, on top of the GOME global ozone field of October 15, 2010 (source Environment Canada).



3.5.2 Nadir ozone profile validation data sources

As described in DARD [RD6], the following measurement data sets are used as reference for validation studies and/or for cross-comparison studies of the nadir ozone profile data products:

- Ground-based ozone profile measurements by balloon-borne electrochemical ozonesondes.
- Ground-based ozone profile measurements by stratospheric ozone lidars.
- Optionally (not done here), where appropriate, satellite measurements of stratospheric ozone profile at higher vertical resolution contributing to the limb CRDP, and also ERBS SAGE-II, UARS HALOE and EOS-Aura MLS.

3.5.2.1 Ozonesonde measurements

In-situ measurements of ozone are carried out regularly by ozonesondes on-board small meteorological balloons launched at numerous sites around the world. They measure the vertical profile of ozone partial pressure with 100 to 150 meter vertical resolution from the ground to the burst point of the balloon, usually between 30 and 35 km. An interfaced radiosonde provides the pressure, temperature and GPS data necessary to geolocate each measurement or to convert the ozone partial pressure to other units. Normalisation factors, if provided, are not applied. Different types of ozonesondes were developed over the years. Those still in use today are based on the electrochemical reaction of ozone with a potassium-iodide sensing solution. Laboratory tests and field campaigns indicate that between the tropopause and about 28 km altitude all sonde types produce consistent results when the standard operating procedures are followed [RD79]. The bias is smaller than $\pm 5\%$ and the precision is about 3%. Above 28 km the bias increases for all sonde types. Below the tropopause, due to lower ozone concentrations, the precision degrades slightly from 3 to 5%, depending on the sonde type. The tropospheric bias also becomes larger, between ± 5 to $\pm 7\%$. Other factors besides ozonesonde type influence the data quality as well. A detailed overview can be found in [RD79]. The present work relies on the ozonesonde data archived by the Network for the Detection of Atmospheric Composition Change (NDACC), Southern Hemisphere Additional Ozonesonde network (SHADOZ; [RD84], [RD85]) and WMO's Global Atmospheric Watch (GAW). Together these three data sources cover 82.5° N to 90.0° S and provide soundings at least once a week at many participating stations. Stations contributing to the present study are highlighted in Figure 3.6.

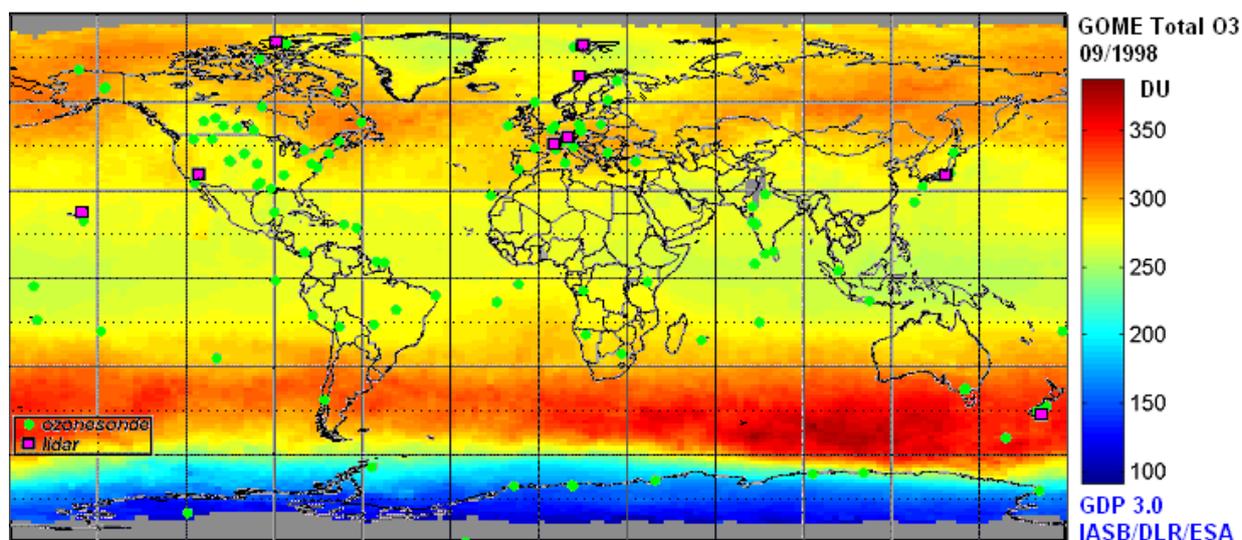


Figure 3.6 - Geographical distribution of ground-based NDACC lidar and GAW ozonesonde stations having archived regularly ozone profile data to the NDACC DHF, SHADOZ archive and/or the WOUDC during the Envisat era, displayed on top of a total ozone map typical of September.



3.5.2.2 Lidar measurements

A differential absorption lidar (DIAL) operates mostly during clear-sky nights, simultaneously emitting two pulsed laser beams at wavelengths with a different ozone absorption cross-section. The backscattered signal is integrated over a few hours to retrieve the vertical distribution of ozone [RD67]. A stratospheric ozone lidar system emits beams at 308 nm and 353 to 355 nm, which makes it sensitive from the tropopause up to about 45 to 50 km altitude with a vertical resolution that declines with altitude from 0.3 to 3 km. The profiles are reported as ozone number densities versus geometric altitude. The DIAL technique is in principle self-calibrating since the ozone profile is retrieved directly from the returned signals without introducing instrumental constants. However, interference by aerosols, signal induced noise and saturation of the data acquisition system can degrade the quality of the measurements. Unreliable measurements can be discarded based on the reported precision, which were shown to be realistic [RD40]. The bias and precision are about $\pm 2\%$ between 20 to 35 km, increasing to ± 5 to $\pm 10\%$ outside this altitude range where the signal-to-noise ratio is smaller [RD45]. The consistency between six ozone lidars in the NDACC network was recently studied using various satellite data sets [RD69]. This study concluded that the different lidar records agree within $\pm 5\%$ of the space-based observations over the range of 20 to 40 km. Data from all stratospheric ozone lidars that have been operational in the NDACC network since the beginning of the 1990s and cover year 2008 are considered. The network covers 80.0° N to 67° S, but most sites are located in the northern hemisphere. Lidar stations contributing to the present study are highlighted in Figure 3.6.

3.5.3 Limb ozone profile validation data sources

The following measurement data sets are used as reference for validation studies and/or for cross-comparison studies of the limb ozone profile data products:

- Ground-based ozone profile measurements by balloon-borne electrochemical ozonesondes.
- Ground-based ozone profile measurements by stratospheric ozone lidars.
- Ground-based ozone profile measurements by ozone microwave radiometers.
- Optionally (not done here), satellite measurements of stratospheric ozone profile by ERBS SAGE-II, UARS HALOE and EOS-Aura MLS.
- Optionally (not done here), satellite ozone profile measurements at lower vertical resolution but with global coverage, by the series of SBUV/2 on NOAA-9/11/14/16/17/18 operational polar satellites and by EOS-Aura OMI.

3.5.3.1 Ozonesonde measurements

Details of the ozonesonde measurement technique, associated uncertainties and contributing stations are given in Section 3.5.2.1.

3.5.3.2 Lidar measurements

Details of the lidar measurement technique, associated uncertainties and contributing stations are given in Section 3.5.2.2.

3.5.3.3 Microwave radiometer measurements

Microwave radiometers (MWR) record the emission of a thermally excited rotational transition at 110 or 142 GHz. Observations are integrated over 1-4 hours and they are carried out continuously during day and night, irrespective of cloud conditions or aerosol load. Vertical profiles of ozone VMR are retrieved on fixed pressure levels between 20-25 and 70 km from the pressure broadening of the integrated line spectra. Ozone VMR can be converted a posteriori into number density using meteorological (re)analyses of pressure and temperature. The total uncertainty of ozone retrievals is estimated at less than 10-15% between 25-50 km and increases to 25% at the profile top and bottom. When compared to ozonesonde and lidar the vertical resolution of MWR is much poorer, about 8-10km in the stratosphere up to 15 km in the mesosphere. On the other hand, the number of measurements is superior, so the co-location criteria can be stricter to reduce uncertainties in the comparison results due to spatiotemporal mismatch. We consider the MWR ozone



profile data uploaded to NDACC data host facility by several stations: in Bern (47.0°N, 7.4°E), Payerne (46.8°N, 7.0°E), Mauna Loa (19.5°N, 155.6°W) and Lauder (45.0°S, 169.7°W).

3.5.4 Error budget of the comparison of atmospheric data

A major objective of quantitative comparisons with reference measurements and modelling results of documented quality is to estimate uncertainties of the validated data product and to check the accuracy of its theoretical uncertainty estimates. However, in fact the systematic and random discrepancies between the validated data set and the validation data set combine uncertainties associated with each individual system, plus uncertainties associated with the selection of data and the methodology of comparison [RD59]. Discrepancies include the effect of the following comparison uncertainties:

- (1) Comparison uncertainties associated with the difference in sampling of atmospheric variability and structures: e.g., geographical mismatch, diurnal cycle effects in the upper stratosphere and mesosphere (USM), assumptions related to the area of representativeness.
- (2) Comparison uncertainties associated with the difference in smoothing of atmospheric variability and structures: e.g., balloon-based in situ measurement at about 150 m vertical resolution by an electrochemical cell, compared with GOME ground pixels of 40 x 320 km² and vertical resolution of 3-8 km.

As far as possible, most comparison uncertainties will be reduced by a cautious design of the selection of data sets to be compared, and by considering that a multivariate analysis of the comparison results taking into account the specifics of the data being compared (modelling data or remote sensing data, atmospheric variability and gradients etc.) might be required and preferred over entirely statistical approaches. For traceability purposes it is essential to document for each validation exercise the selection method applied to the data sets (temporal and spatial co-location criteria, how differences in vertical and horizontal smoothing are handled etc.).

Although essential if a rigorous metrological approach is to be adopted, the derivation of a complete error budget for each comparison is still a matter of research at the time being and it falls partly beyond the scope of the Ozone_cci+ project. In Section 7, a first proof-of-concept is elaborated in which the agreement between satellite (S5P) and ground-based reference measurements of the total column of ozone is quantified in terms of their combined ex-ante error budgets. While agreement within the ex-ante uncertainties is found for this particular case, this can not be assumed to hold also for products where the comparison method is more complex and introduces significant additional error sources, e.g. in the case of vertical profiles derived from nadir measurements. Validation teams as well as EOSTs are aware that neglecting uncertainties linked to the comparison method can spoil the value of the comparison and yield erroneous conclusions on the quality of the compared data product. With this disclaimer an awareness is transmitted to the reader of Ozone_cci+ Validation Reports for proper use of the validation results and, if fine, of the CCI ozone CRDP.



4 Validation of Total Ozone Data Products

4.1 Scope and generalities

The Ozone_cci+ CRDP total ozone datasets include Level-2 data records acquired by ERS-2 GOME, Envisat SCIAMACHY, Aura OMI, GOME-2 Metop-A, Metop-B and Metop-C, and TROPOMI/S5P as well as the Level-3 merged data product built upon those seven individual datasets. SNPP OMPS is also processed with the GODFIT v4 algorithm and validated here. Within Ozone_cci+, TROPOMI/S5P and GOME-2/Metop-C are recently introduced for the extension of the CRDP datasets, therefore, their validation results are separately shown in Sections 4.2.1 and 4.2.2. Additionally, the IASI Metop-A, Metop-B and Metop-C Level-2 total ozone products retrieved by the FORLI-O3 v20191122 processing algorithm, are also validated.

Therefore, this section starts with the detailed validation results of the Level-2 data sets (Sections 4.2 and 4.4), analysed using the same methodology and it continues with the validation of the Level-3 merged dataset (Section 4.3). While the objectives of the Level-2 validation are classical (determination of the systematic bias, dependences on SZA and clouds, etc.), the purpose of the Level-3 validation study is to demonstrate that compared to the traditional Level-2 validation results, no spurious features appear, or no new features in general. Finally, the validation of the MSR dataset is presented (Section 4.5).

To prepare the ground-based data set, we have investigated the quality of the total ozone values of each station and instrument that deposited data at WOUDC and NDACC after 1995. The selection methodology and associated criteria have been discussed in detail in [RD57, RD26, RD58, RD41 and RD42]. We offer here a brief summary for completeness. For each ground-based station a series of statistics and plots are performed. Daily coincidences of the satellite pixel's central latitude and longitude falling within a 150km radius (10 km for TROPOMI/S5P) of the ground station are identified and used for the creation of monthly, seasonal and yearly time series and scatter plots. The percentage of the relative differences between ground and satellite TOC is used as the comparative tool for the validation. The statistics are then typically performed on a zonal average, on a hemispheric average and on a global average, always keeping the two types of ground-based instruments separate and using only direct sun observations, as they are deemed to be the most reliable.

The MSR Level-4 data set is created in a 2-step process, in which the satellite data sets are bias corrected before being assimilated. This bias correction relies on a comparison to Brewer and Dobson data, and as such these ground-based measurements can no longer serve as independent reference data. For that reason, the validation is done solely with the NDACC ZSL-DOAS data. Further particulars about filtering, co-location and averaging are described in Section 4.5.

4.2 Level-2 total ozone retrieved with GODFIT v4

The Level-2 validation results summarised hereafter are an update of the Ozone_cci GODFIT v4 satellite TOCs validation work, which has been presented in Garane et al. (2017) [RD38]. Except for the extended (w.r.t. RD38) data sets of GOME2A, GOME2B and OMI, new sensors are added for the purposes of this report, according to the Ozone_cci+ validation plan. In the following sections, the validation results of TROPOMI/S5P and GOME2-MetopC are also presented, since the two sensors are integrated in the GTO-ECV dataset. It should be noted that the TROPOMI/S5P TOC retrieval is performed using the OFFL algorithm [RD75], which relies on GODFIT v4 algorithm. The previous version of the GODFIT algorithm is described fully in the Ozone_cci GODFIT v3 retrieval algorithm paper by Lerot et al. (2014) [RD61] and its validation details by Koukoulis et al., in the Ozone_cci GODFIT v3 validation paper [RD52]. The updates of the retrieval algorithm applied in the latest version 4 are presented in the Ozone_cci+ ATBD [RD4]. Table 4.1 shows the time span of all Level 2 datasets that will be used here.



Table 4.1 - The time span of the available Level-2 datasets processed with GODFIT v4.

Sensor	Algorithm	Time span	
GOME/ERS-2	GODFIT v4	06/1995 – 06/2011	16 years
SCIAMACHY/Envisat		08/2002 – 04/2012	10 years
OMI/Aura		10/2004 – 12/2021	17 years
GOME-2/MetopA		01/2007 – 04/2021	14 years
SNPP/OMPS		01/2012 – 12/2021	10 years
GOME-2/MetopB		01/2013 – 11/2021	9 years
GOME-2/MetopC		02/2019 – 12/2021	3 years
TROPOMI/S5P	OFFL	11/2017 – 12/2021	4 years

4.2.1 TROPOMI/S5P Level-2 total ozone

The TROPOMI/S5P total ozone column product retrieved by the OFFL algorithm [RD39] is integrated to the GTO-ECV products in the frame of the Ozone_cci+ project. This total ozone product has been officially validated by Garane et al. (2019) [RD39]. According to the validation conclusions, the TROPOMI OFFL TOC has a mean relative bias with respect to ground-based measurements up to 1%, well within the product's requirements. The aforementioned results were based on the analysis of just the first year of TROPOMI measurements (Nov 2017 - Nov 2018). To date, 4 years of data are available, and the product is continuously validated (see <https://mpc-vdaf.tropomi.eu/index.php/total-ozone?start=4>). In the following, some updated validation results are presented.

In Figure 4.1, the scatter plots and histograms of the overall percentage differences between co-located TROPOMI and ground-based TOC measurements are shown. Panels (a) and (c) show the Dobson comparisons; the Brewer comparisons (for the Northern Hemisphere only) are used for the plots in panels (b) and (d). The correlation coefficient in both cases is very satisfactory, above 0.98, proving the very good agreement between satellite and ground truth, while the mean percentage relative bias is below 0.9 %, within the requirements. Furthermore, the distribution of the percentage differences around their mean is normal in both cases. The higher standard deviation in the Dobson comparisons is due to the limited number of stations located at the Southern Hemisphere, compared to the Northern Hemisphere's station density. This is also depicted in Figure 4.2 (panel a), where the mean percentage difference of each station is presented, with respect to the stations' latitude (Dobson-panel a and Brewer-panel b). Additionally, some of these SH stations are not as quality assured as others, but they cannot be excluded from the comparison dataset because, in some cases, they are the only source of ground-based information in the area.

Also in Figure 4.2, the time-series in terms of monthly means of the co-locations are shown for both Hemispheres for Dobson (panels c and e) and NH only for the Brewer (panel d) comparisons. In July 2021 the OFFL TOC product UPAS processor was changed from v1 to v2, which caused the increase in bias seen in the time-series. Nevertheless, the time-series of the monthly means is always within the products' requirements, showing its good quality and stability in time.

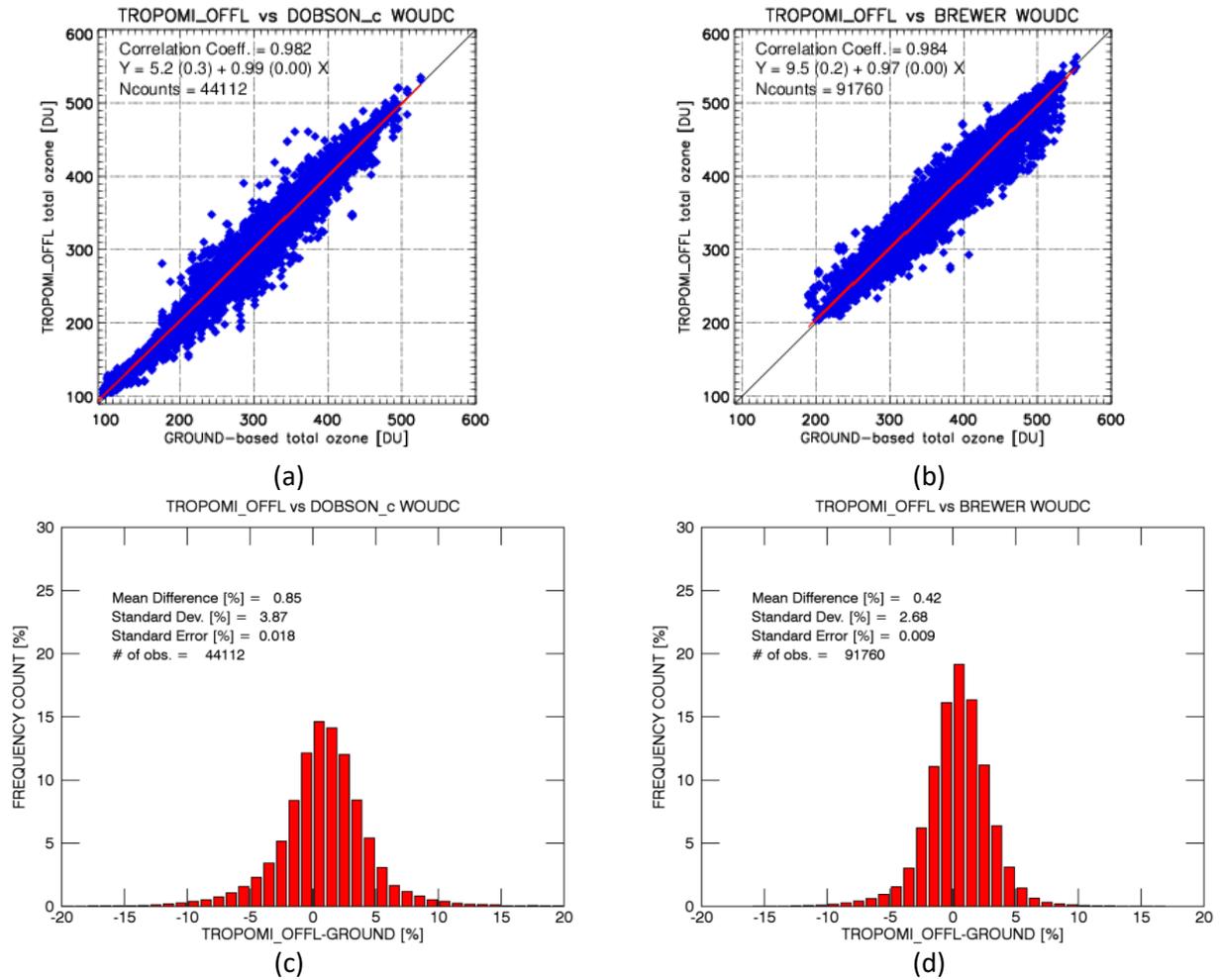


Figure 4.1 - The scatter plots (panels a and b) and histograms (panels c and d) of the TROPOMI/S5P OFFL TOC w.r.t Brewer (right) and Dobson (left) ground-based TOC measurements.

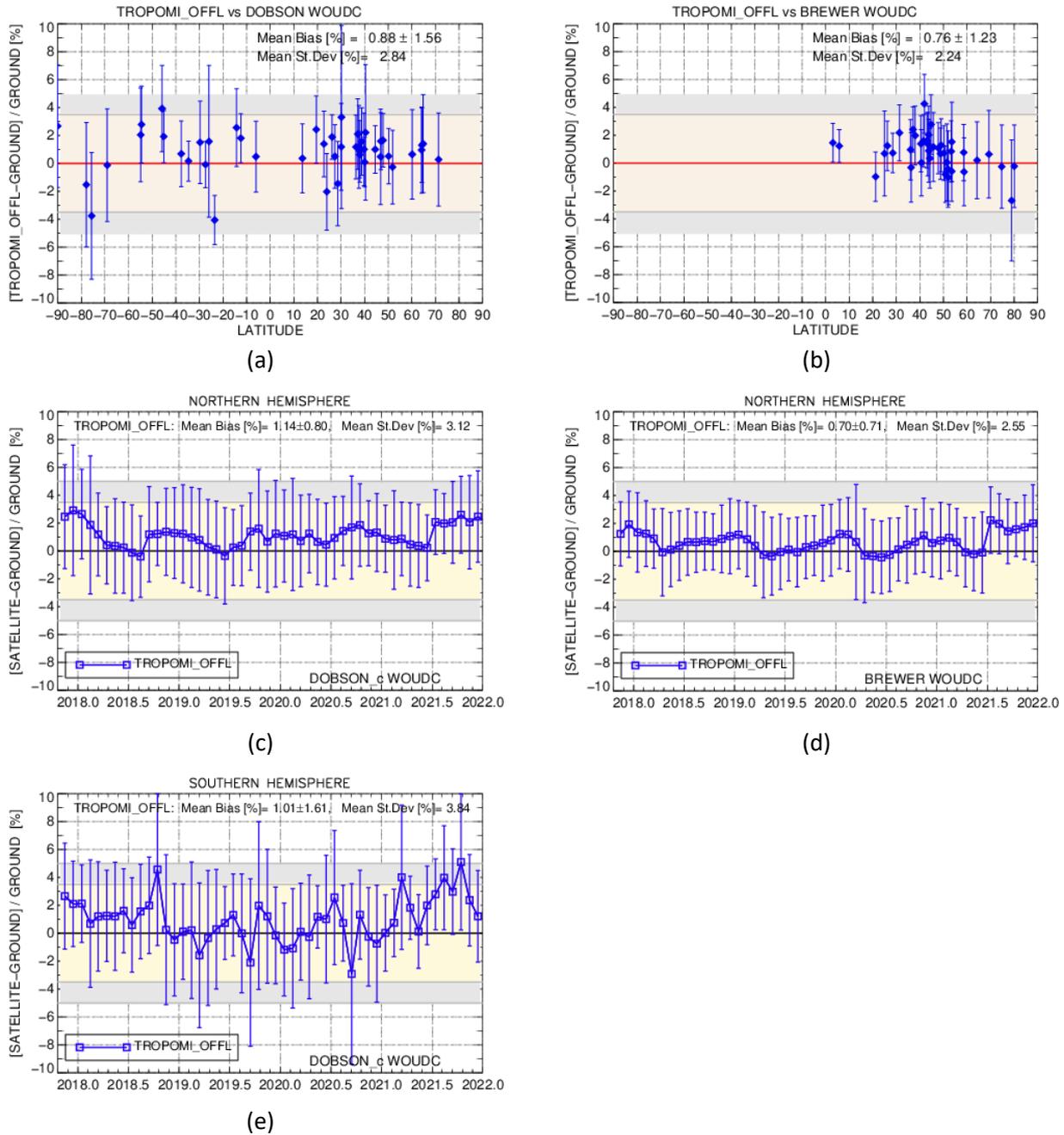


Figure 4.2 – Panels a & b: The mean percentage difference per station between the TROPOMI/S5P OFFL TOC and the Dobson (left) and Brewer (right) networks. **Panels c, d & e:** the hemispherical monthly mean time-series of the co-locations for Dobson (left column) and Brewer (right column, NH only) comparisons.



4.2.2 GOME-2C Level-2 total ozone

Metop-C is the third and last satellite of the Metop series that forms the space segment of the EUMETSAT Polar System (EPS) and it was launched in November 2018 from the European Space Port in French Guyana. The total ozone measurements from the GOME-2 instrument onboard Metop-C (GOME-2C) are also recently integrated to the ECV data of the Ozone_cci+ project. The GOME-2C TOC data retrieved by the GODFIT v4 algorithm cover the period from February 2019 to December 2021, almost 3 years. In the following, this new product is validated against ground-based instruments and their co-located measurements.

In Figure 4.3 the overall statistics of the satellite and ground-based instruments (Dobson to the left and Brewer network to the right) co-locations are shown. In panels a and b, the scatter plots show that the correlation coefficient is above 0.97, indicating already the very good agreement. The histograms in panels c and d, show the near-perfect normal distribution of the percentage differences. The overall mean relative difference between the satellite sensor and ground-based instruments is positive, indicating that GOME-2C reports higher TOCs than the ground-truth, and the mean difference is $\sim 1.3\%$.

Like Figure 4.2, Figure 4.4 shows the mean percentage difference per station for the Dobson (panel a) and Brewer (panel b) networks and the monthly mean time-series of the co-locations (panels c, d and e), also for Dobson (left) and Brewer (right – NH only) comparisons. As seen before, the same pattern in the latitudinal distribution of the percentage differences occurs, following mainly the ground-based measurements features. The monthly mean time-series plots show a very stable in time agreement between the satellite and ground-based data, especially for the Brewer comparisons in the NH.

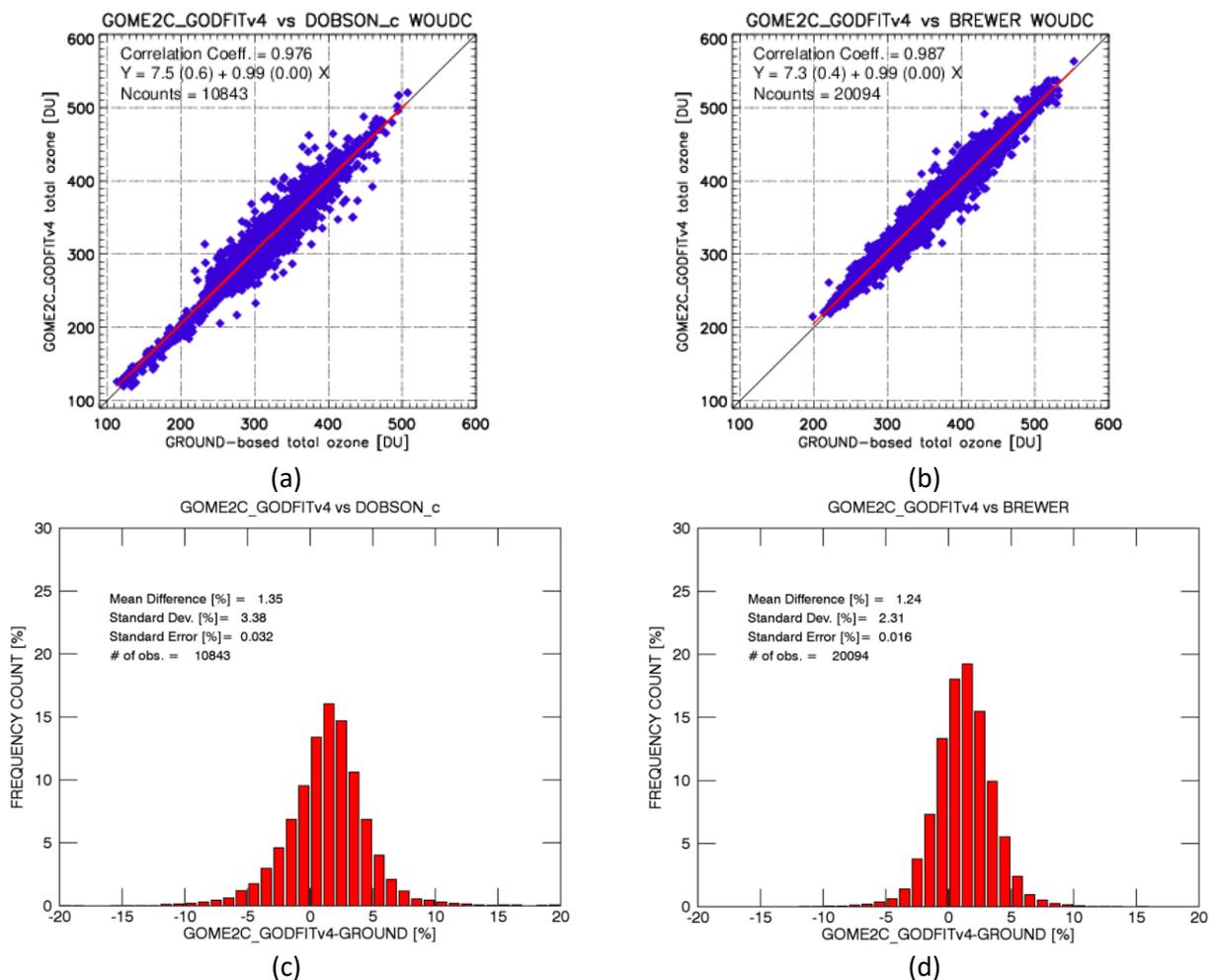


Figure 4.3 - The scatter plots (panels a & b) and histograms (panels c & d) of the GOME-2/Metop-C TOC w.r.t Brewer (right) and Dobson (left) ground-based TOC measurements.

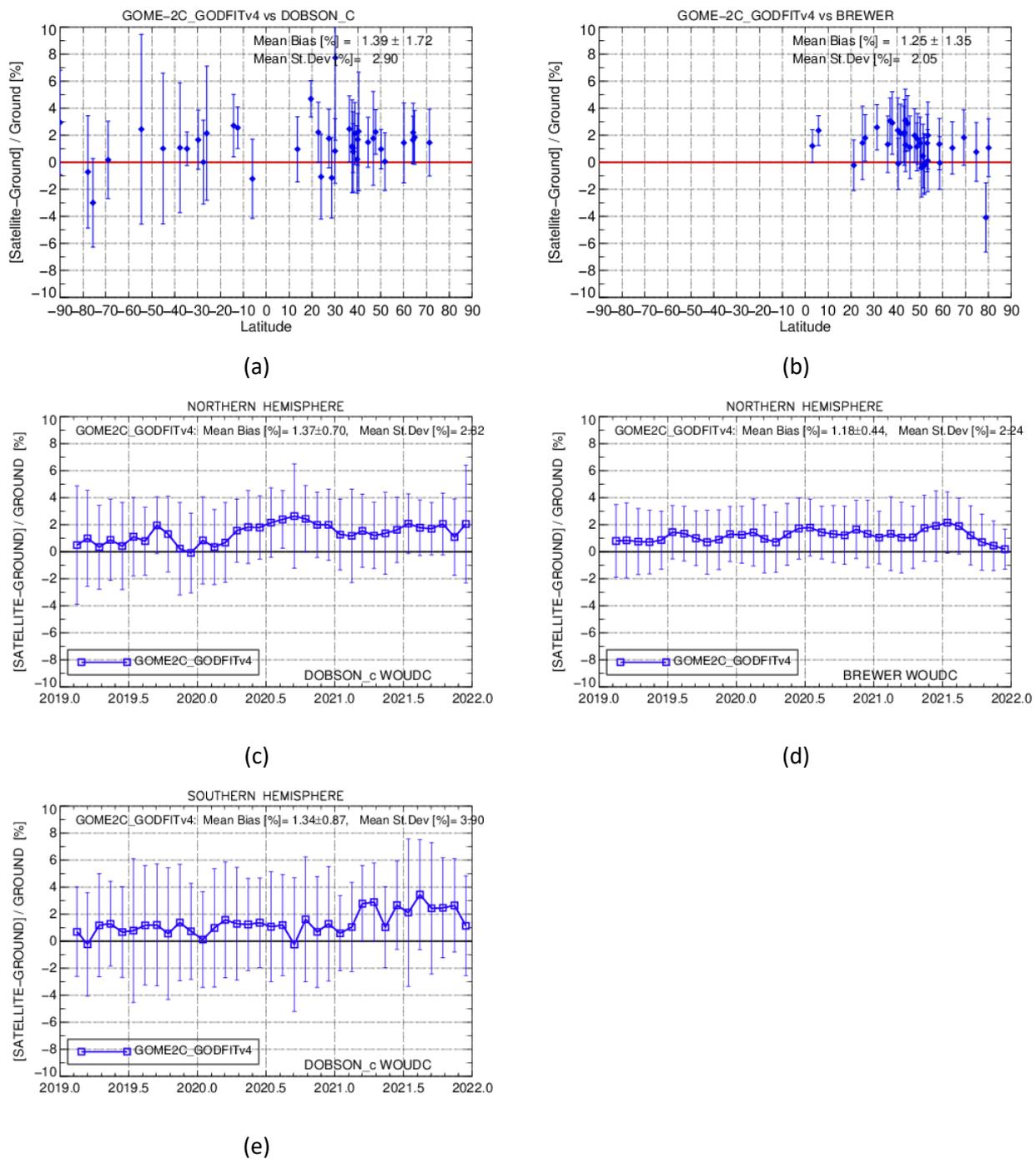


Figure 4.4 – Panels a & b: the mean percentage difference per station between the GOME-2/Metop-C TOC and the Dobson (left) and Brewer (right) networks. **Panels c, d & e:** the hemispherical monthly mean time-series of the co-locations for Dobson (left column) and for the Brewer (right column, NH only) comparisons.



4.2.3 Overview of the Level-2 systematic bias and its variations

In Figure 4.5, the long-term behaviour of each of the eight investigated instruments is examined per latitude belt, as a contour time series plot. The colours depict the percentage difference between satellite and Dobson TOCs on a global scale (pure white colour denotes area of no comparisons, while the 0% level is given as a light beige colour). The respective Brewer comparisons are not shown since they only cover the Northern Hemisphere.

Very similar pictures are presented by the GOME, OMI, GOME-2A and -2B TOCs, with (as expected) higher positive discrepancies above the polar circles than for mid-latitudes and the tropics. The negative discrepancies in the southern polar area is a common feature for SCIAMACHY and OMPS. Finally, TROPOMI and GOME-2C show a very similar distribution of the differences in time and latitude.

The higher differences in the polar regions is attributed both to the satellite instrument/algorithm capabilities as well as to the increase in systematic errors in the ground-based instrumentations, the latter being due to observations at high solar zenith angles and the associated stray light effect increasing in magnitude and importance. Furthermore, there is less availability of quality-assured ground-based measurements in the more inaccessible regions, mainly in the South. The mean difference falls around the 0-2% levels, with some peaks around ± 4 % for the high latitudes.

4.2.4 Inter-sensor stability

In Figure 4.6 the inter-sensor stability is examined for the eight TOC records via a monthly mean time series of the differences between each sensor and the co-located Dobson observations for the Northern Hemisphere (panel a) and the Southern Hemisphere (panel c). The respective time-series for the Brewer co-locations (NH only) are shown in panel b. The excellent agreement in the Northern Hemisphere for both networks, Dobson and Brewer, is undoubtable. In the Southern Hemisphere, OMI, GOME-2A, GOME-2B, GOME-2C and TROPOMI are in very good agreement, while SCIAMACHY appears to deviate by $\sim 1\%$ in terms of mean bias. Nevertheless, when the co-locations southwards 50° S are excluded (not shown here), SCIAMACHY comes to a very good agreement with the other sensors. In the panel c, the ERS-2 GOME (black line) is shown only up to year 2004 due to the change in the SH coverage caused by an instrumental failure which was introducing an unphysical high scatter in the monthly mean calculations.

To better investigate the consistency of the newer sensors, i.e. TROPOMI and GOME-2C, to those that are already part of the GTO-ECV, the monthly mean time series of the percentage differences since 2018 is shown in Figure 4.7. In the Northern Hemisphere, the agreement between all sensors is excellent, within $\pm 1-2$ %. In the Southern Hemisphere, as it was expected, the time-series are much noisier due to the contribution of the high latitude stations and their respective measurements at high solar zenith angles. Excluding stations southwards 50° S results also to an agreement of 1-2 % between all sensors.

Table 4.10 summarizes the mean bias and its standard deviation for all eight sensors and their comparisons to Dobson and Brewer ground-based measurements, with respect to various averaging parameters such as latitude and solar zenith angle, which will be studied in the following sub-sections. Overall, the hemispherical mean relative biases of the eight sensors are within 0.5 and 1.5 % and their variability is within $\pm 3\%$, both quantities within the Requirements.

4.2.5 Seasonality

In Figure 4.8 the seasonal dependency of the percentage differences of all satellite sensors with respect to the ground-based measurements, is shown. In panels a and c, the Dobson comparisons for the Northern (panel a) and the Southern Hemisphere (panel c) are shown. To the right, the seasonality of the Brewer comparisons is depicted (panel b).

All eight sensors agree very well in terms of seasonality, especially in the NH where the number of ground-based stations, reporting TOC measurements regularly, is higher and more homogeneously distributed. The mean seasonal variation is 0.7 - 2% peak-to-peak for the Brewer comparisons in the NH. Since the well-known dependency of the Dobson measurements on effective temperature was accounted for via the application



of the respective post-correction (Section 3.5.1), their peak-to-peak difference results almost equal to the Brewer seasonality, 0.6 – 1.7%.

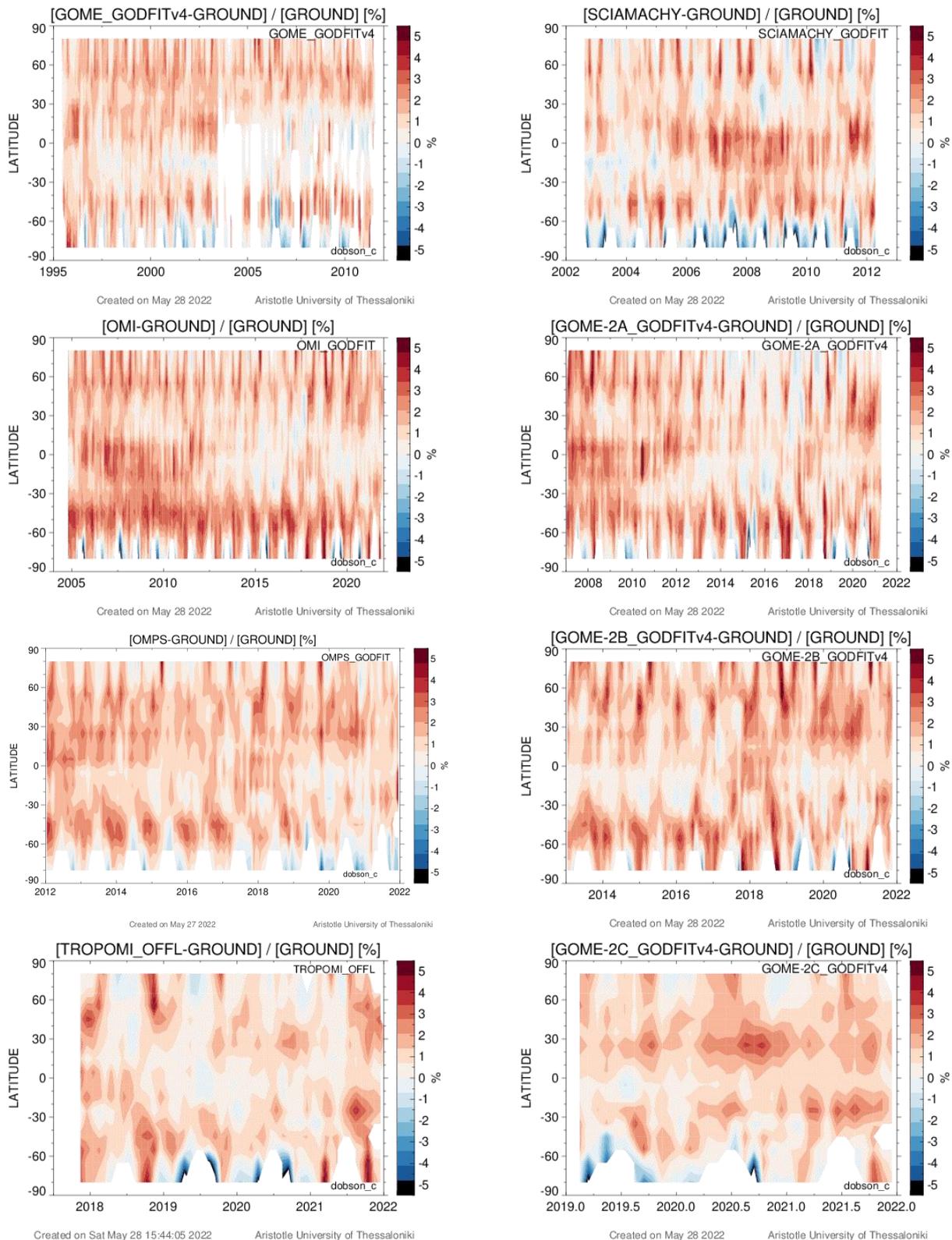


Figure 4.5 - Latitude-time evolution of the percent relative differences, from top to bottom and left to right, between GOME, SCIAMACHY, OMI, GOME-2A, GOME-2B, OMPS, TROPOMI and GOME-2C satellite total ozone data and Dobson network measurements.

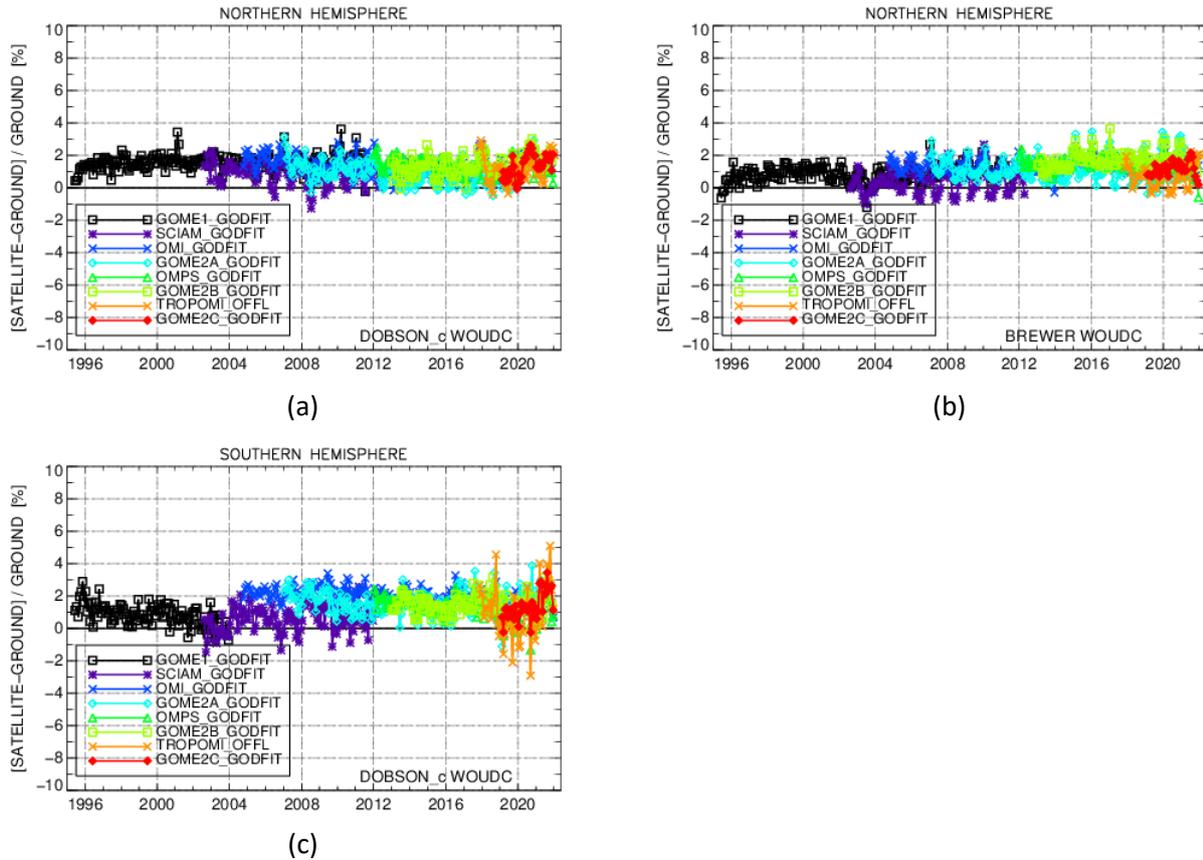


Figure 4.6 - The inter-sensor stability for the entire 26-year time span for all eight instruments against the Dobson (panel a - NH, panel c - SH) and the Brewer network (panel b - NH only).

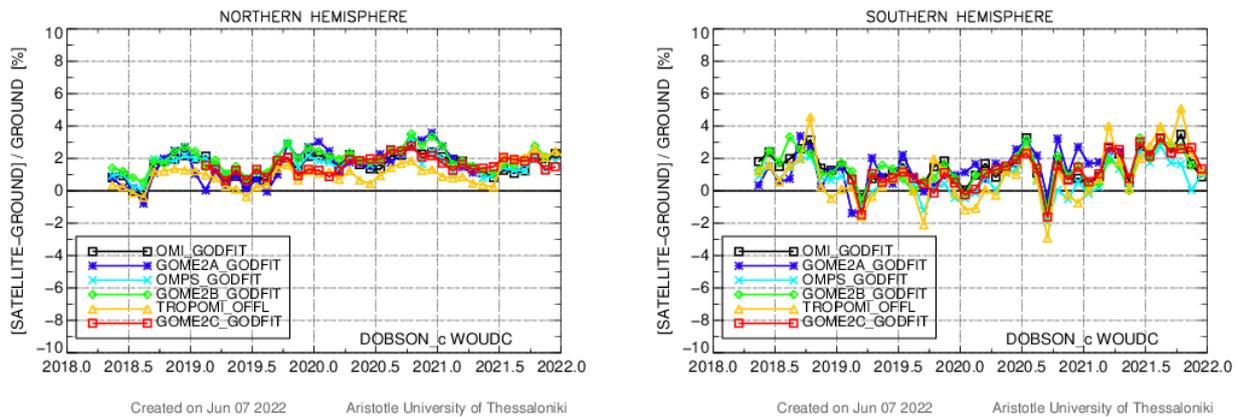


Figure 4.7 - As in Figure 4.6, but for the time span of TROPOMI and GOME-2C operation (2018 onwards).

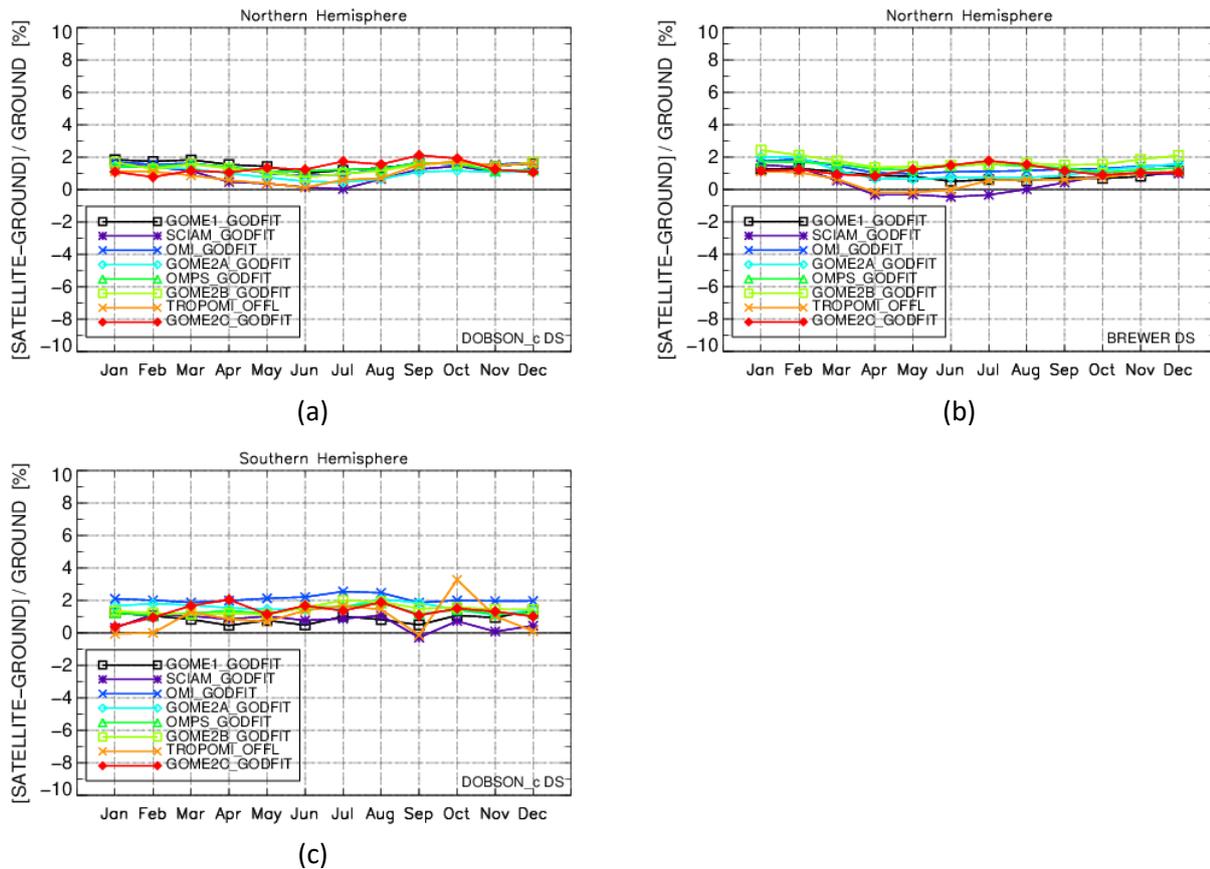


Figure 4.8 - The seasonal dependency of the percentage differences between the eight satellite sensors and the ground-based measurements from Dobson (panels a-NH and c– SH) and Brewer (panel b, NH only) instruments.

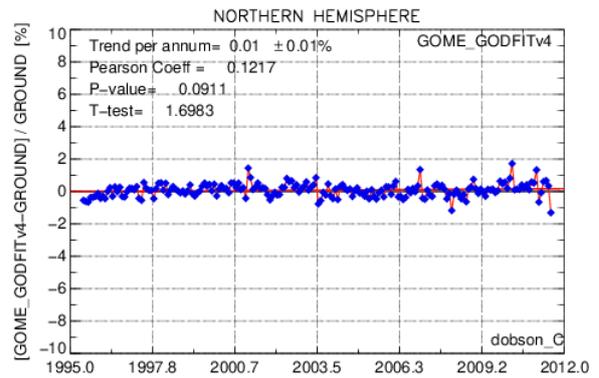
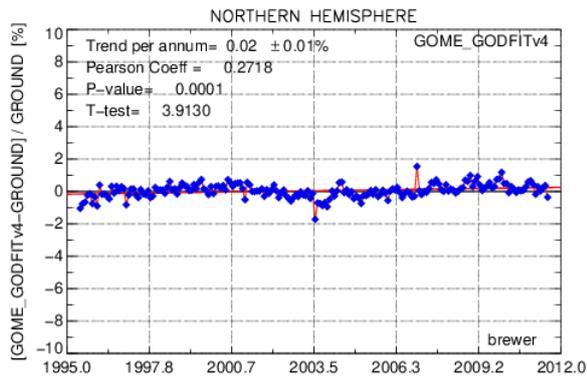
4.2.6 Long-term stability

The following plots (Figure 4.9 to Figure 4.11) show the long-term evolution of the deseasonalized relative differences between GODFIT v4 and ground-based network total ozone data, averaged over the Northern Hemisphere. The GOME, SCIAMACHY, OMI, GOME-2A, GOME-2B, OMPS, TROPOMI and GOME-2C NH time series for the Brewer (left panels) and the Dobson co-locations (right panels), are shown. Note that in all graphs the calculated drifts of the monthly mean time series are shown. The GOME-2C dataset spans only 3 years of available data, therefore even though its long-term stability is shown here, the study cannot be considered statistically significant.

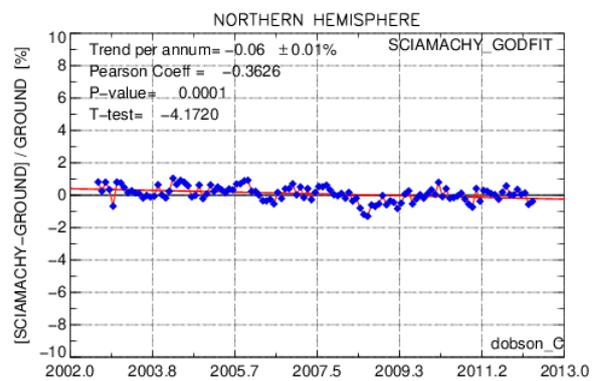
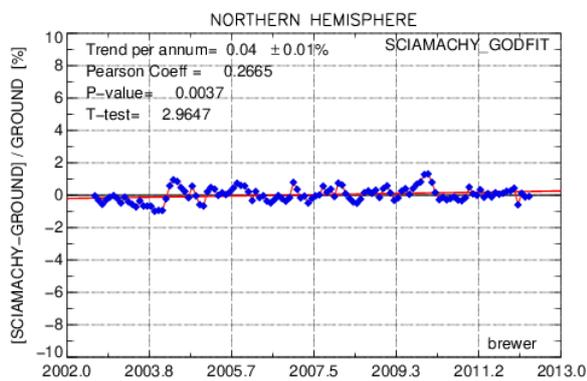
The decadal drifts per decade for all the available sensors are summarized in Table 4.10, where the statistically significant drifts are noted with bold characters. According to this statistical analysis, the reported drifts resulting from the Northern Hemisphere co-locations are very small and always less than 1%/decade, well within the requirements of a long-term stability better than 1-3% / decade. For some sensors, they are not even statistically significant. The GOME-2C drifts are higher, 1.5-2 %/decade, but as mentioned above this should be considered with caution. Additionally, it should be noted that the decadal stability of the satellite data used in this study is close to the stability offered by the ground-based networks.



GOME (GODFIT v4)



SCIAMACHY (GODFIT v4)



OMI (GODFIT v4)

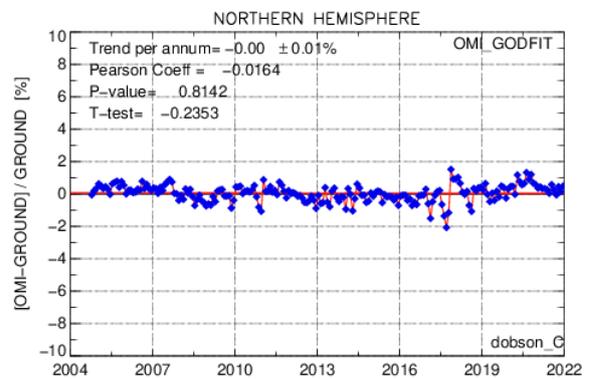
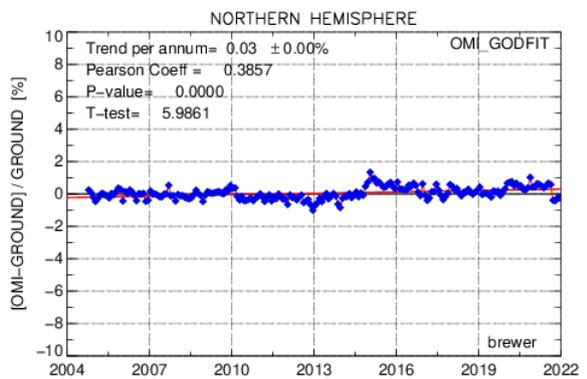
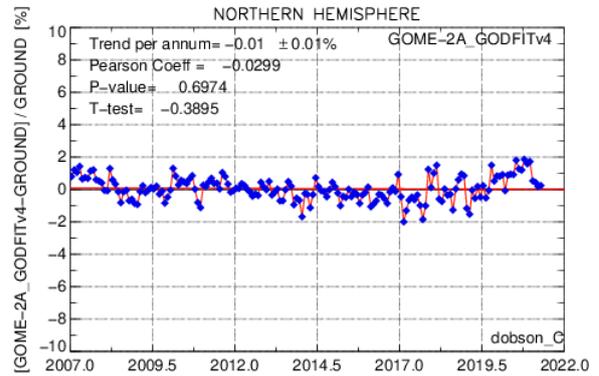
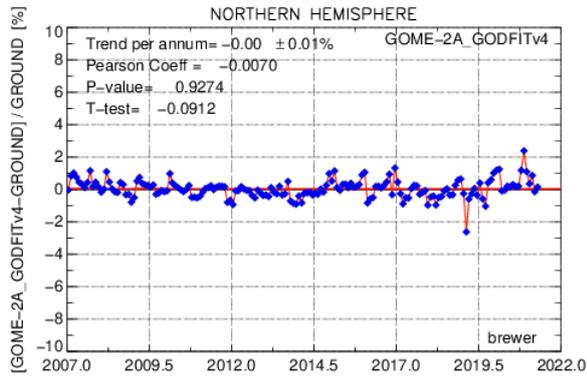


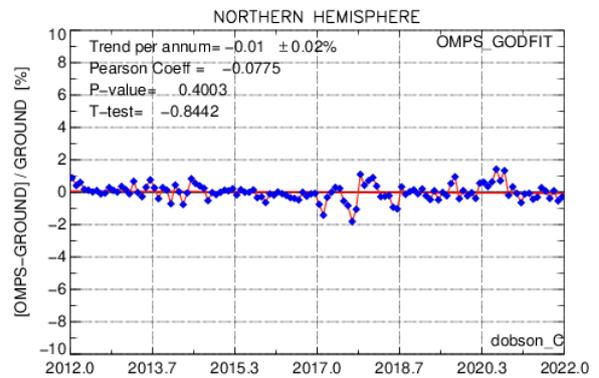
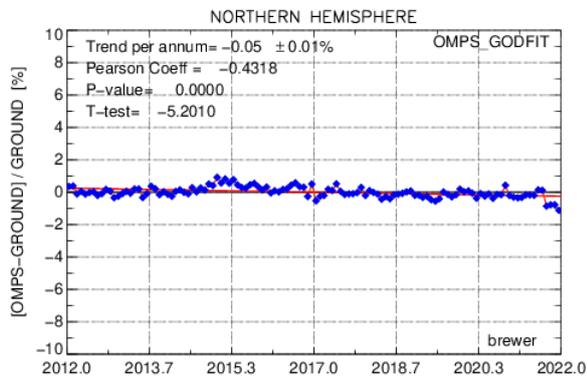
Figure 4.9 - Long-term drift of the deseasonalized percentage relative differences between total ozone data measured by GOME (first row), SCIAMACHY (second row) and OMI (third row) and the Northern Hemisphere Brewer (left column) and Dobson network (right column).



GOME-2A (GODFIT v4)



OMPS (GODFIT v4)



GOME-2B (GODFIT v4)

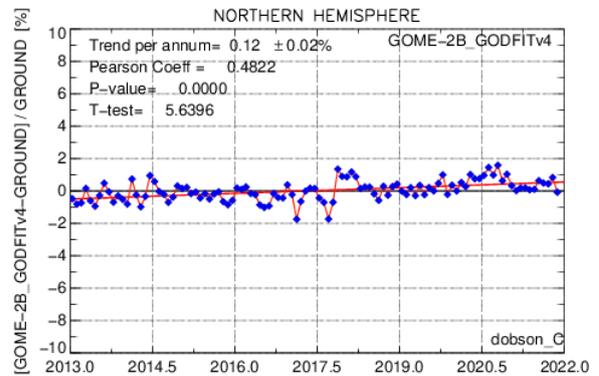
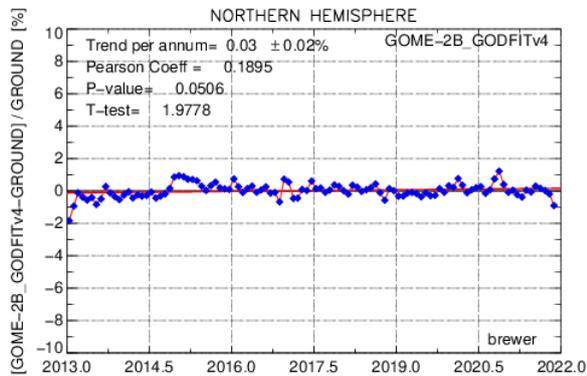
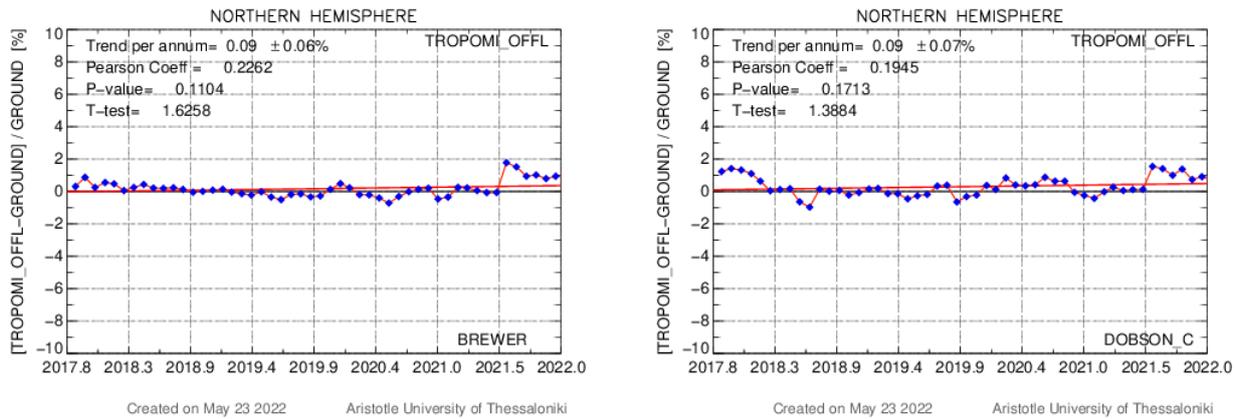


Figure 4.10 – As in Figure 4.9 for GOME-2A (first row), OMPS (second row) and GOME-2B (third row).



TROPOMI S5P OFFL



GOME-2C (GODFIT v4)

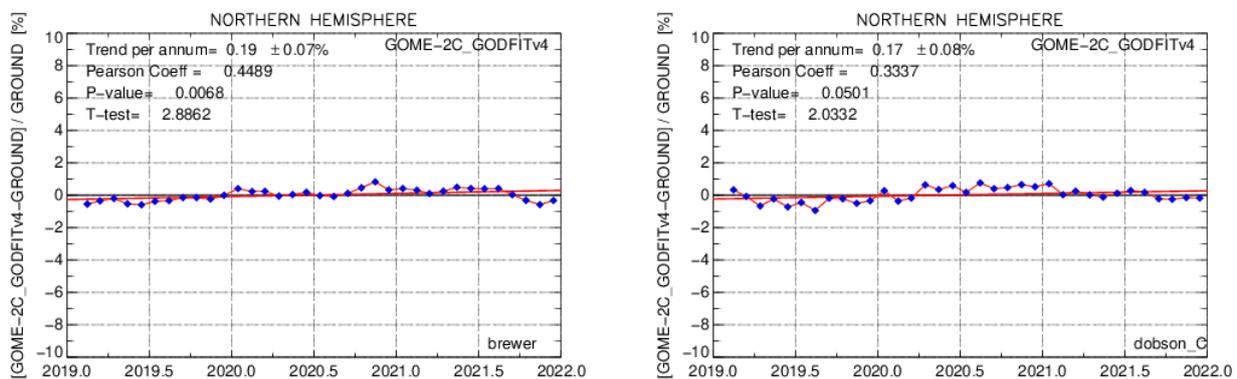


Figure 4.11 - As in Figure 4.9 for TROPOMI/S5P (OFFL) (upper row) and GOME-2C (bottom row).

4.2.7 Dependence on influence quantities

In this section, the dependence of the comparisons between satellite and ground-based total ozone observations on various influence quantities associated with the satellite TOC retrieval, will be investigated.

In Figure 4.12, the meridian (panels a and b) and solar zenith angle (panels c and d) dependency of the differences is depicted as line plots on a global scale for the Dobson (left column) and for the Northern Hemisphere Brewer instruments (right column).

All eight instruments follow exactly the same latitudinal patterns, which are mostly originating from the choice of ground-based instruments as background TOC truth. In detail:

- The Dobson meridian comparisons (panel a): for the latitude belt 70°S to 80°S, the sensors are split in two groups regarding their statistics:
 - SCIAMACHY, OMPS, TROPOMI and GOME-2C have negative mean biases of about -2 to -3%.
 - GOME, GOME-2A, GOME-2B and OMI have higher mean biases ~ -0.5 to 0%

Northwards 40°S, where the station density is better, the difference among the satellite instruments is 1-2%, which is reasonable considering the fact that the latitudinal averaging of each sensor is done using its respective time period of operation.

- The Brewer NH comparisons (panel b): the same difference of 1-2% between all instruments is seen for most latitude belts.

In the SZA dependency plots, the inter-sensor consistency is found to be very good for angles up to 70° or 75°, above which it is impossible to separate the errors introduced by the satellite as well as the ground-



based instrumentation and algorithm capabilities. The agreement is better for the Dobson co-locations for low SZAs, due to the divergence of SCIAMACHY and GOME-2C from the other sensors seen in the Brewer comparisons that result from Northern Hemisphere co-locations only. The statistics (mean percentage differences and standard deviation) of the two averaging methods (latitudinal and for SZA) for all datasets are also summarized in Table 4.10.

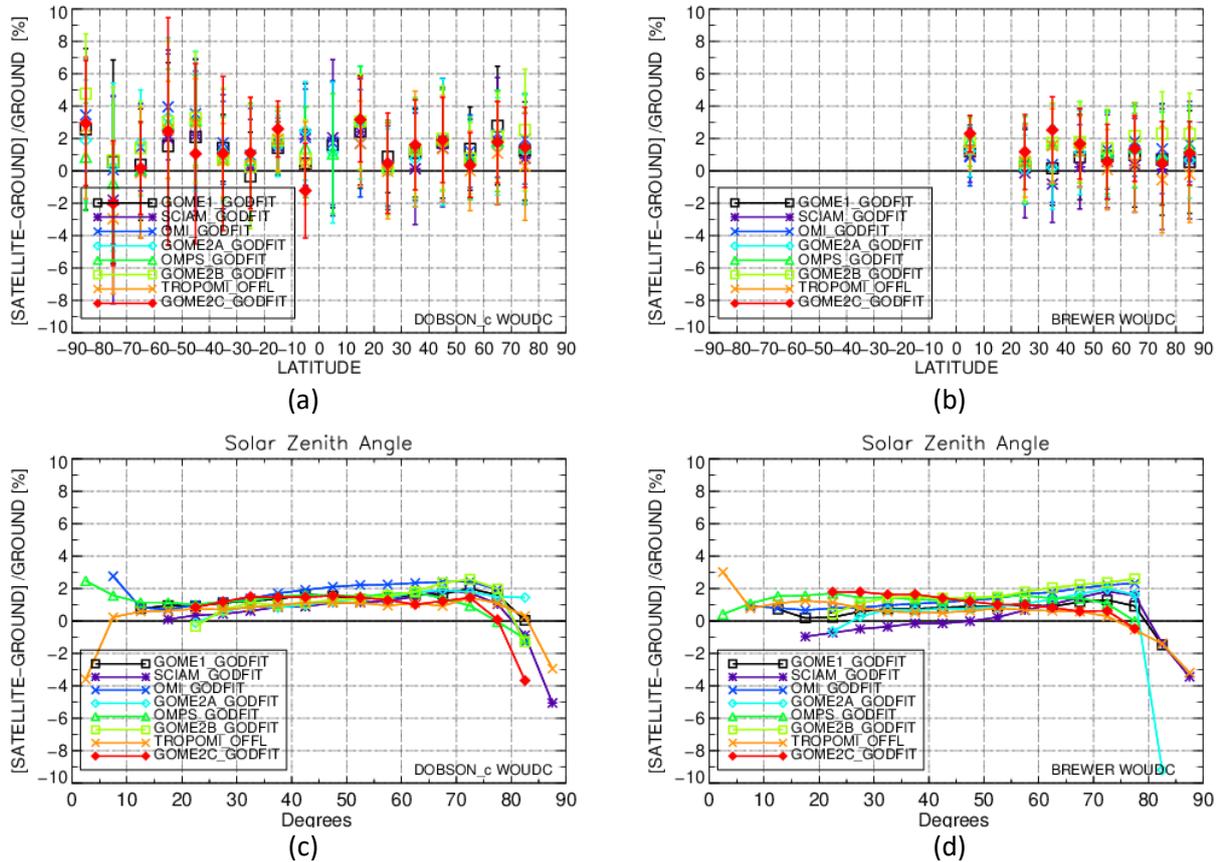


Figure 4.12 - Variation of the percentage relative difference between total ozone data measured by the eight satellite sensors and by ground-based networks (Dobson on the left, Brewer on the right), as a function of latitude (upper row) and the solar zenith angle of the satellite measurement (lower row).

Figure 4.13 shows the dependence of the relative percentage difference between satellite and ground-based Brewer network data on fractional cloud cover, over the Northern Hemisphere. No dependence at all can be detected at any type of cloudiness condition, with the SCIAMACHY and TROPOMI average difference on the 0.3% line and the GOME-2B difference just below the 2% line.

Figure 4.14, left panel, shows the dependence of the relative percentage differences between satellite and ground-based Brewer network data over the Northern Hemisphere, on the effective temperature. The dependence is smooth and remains within the $\pm 1\%$ level for all instruments, especially for co-locations in the temperature range 210 - 235 degrees Kelvin, outside which the features that appear are related mainly to the small number of co-locations for those temperature bins. The sole exception is GOME-2B which shows a small dependency above 230 degrees Kelvin, where the mean deviation increases from 2% to 3% at 237.5 degrees Kelvin. Additionally, GOME-2C has a very limited number of co-locations characterized by effective temperature above 230 degrees Kelvin, which explains its deviation from the other sensors in that particular temperature range. Overall, the GOME-2C differences increase with effective temperature by 3%. As for TROPOMI/SP, it is biased low by up to $\sim -6.5\%$ for a few extreme temperatures, below 215 degrees Kelvin.



Figure 4.14, right panel, shows the respective dependence of the differences on the ghost ozone parameter of the GODFIT v4 algorithm. Except for the bias of each sensor with respect to ground-based measurements, no dependencies on ghost ozone are seen below 12 D.U. For the few co-locations with ghost ozone values greater than 12 D.U., OMPS has a negative bias (~-1%), while the rest of the sensors, mainly GOME-2B, show an increase by up to 3%.

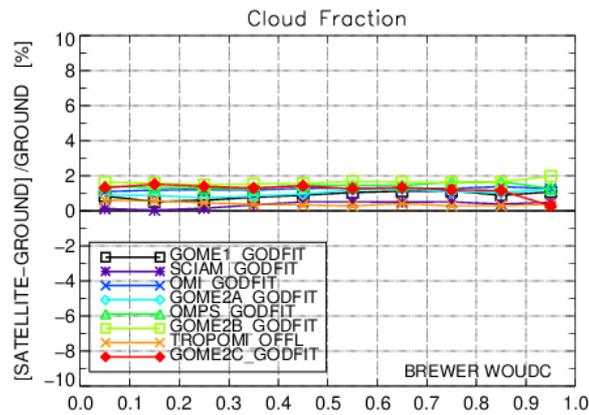


Figure 4.13 - Percent relative difference between satellite GODFIT v4 total ozone data and ground-based Brewer network data over the Northern Hemisphere, as a function of the satellite fractional cloud cover.

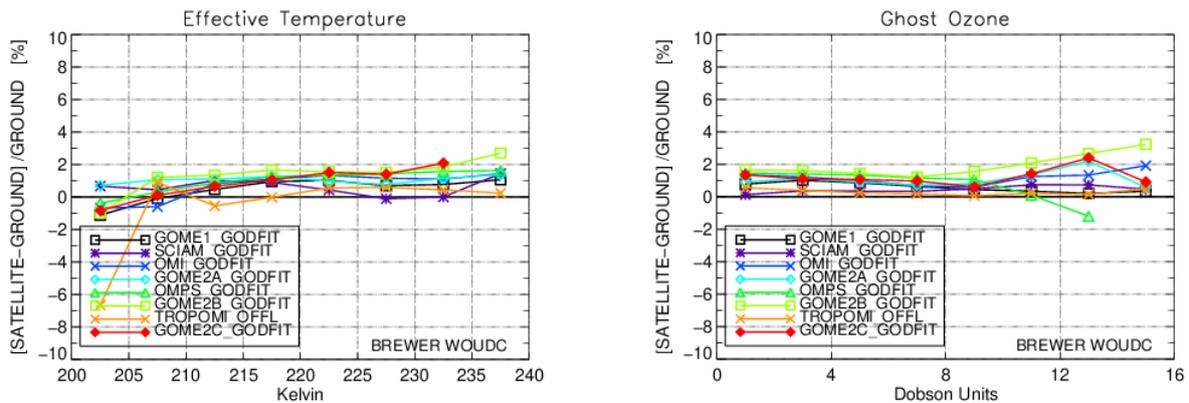


Figure 4.14 – Relative percentage differences between satellite GODFIT v4 total ozone data and ground-based Brewer network data over the Northern Hemisphere, as a function of the effective temperature (left panel) and the ghost ozone (right panel) parameters associated with the satellite TOC retrieval.

4.2.8 Summary and compliance with user requirements

In the above, the Essential Climate Variable (ECV) Climate Research Data Package Total Ozone Column (CRDP TOC), refined and updated via the European Space Agency's *Climate Change Initiative plus*, is presented and validated against independent ground-based TOC observations. Level-2 TOCs, produced by the GODFIT v4 algorithm as applied to the GOME/ERS-2, OMI/Aura, SCIAMACHY/Envisat, SNPP/OMPS, GOME-2/Metop-A, -Metop-B, -Metop-C and TROPOMI/S5P observations, form the basis for a 27-year long consistent, smooth and homogeneous CRDP. Detailed quality control and assurance against specific requirements from the international climate-chemistry modelling community showed that the product more than meets the official User Requirements, i.e., that the stability of the TOC measurements has to be between 1 and 3% per decade, that the radiative forcing introduced by the evolution of the ozone layer has to be less than 2% and that the short-term variability has to be less than 3%. In detail:



- the individual Level-2 data sets show excellent inter-sensor consistency with mean differences within 1.5% at moderate latitudes (+/-50°);
- the mean bias between GODFIT v4 satellite and Brewer and Dobson reported TOCs is well within 0.5 to 1.5 % for all sensors;
- the drift per decade spans between -0.5 to +1.2 %, depending on the sensor and for most of them they are all statistically significant in a 90-99% level.
- The peak-to-peak seasonality ranges between ~0.6 and 1 % for GOME, OMI, GOME-2A, GOME-2B, GOME-2C and OMPS. For SCIAMACHY and TROPOMI/S5P the seasonality is higher, up to ~2%.

Key validation results for the GODFIT v4 and TROPOMI OFFL data records are summarised in Table 4.2 to Table 4.9, respectively. Those tables also reproduce data quality criteria established by the Climate Research Group (CRG) in the Ozone_cci+ User Requirement Document [RD8]. The level of compliance of the CRDP GODFIT v4 datasets with these user requirements is highlighted with a colour code:

- **green** indicates ascertained compliance with requirements from all contributing users;
- **yellow** indicates compliance with requirements from some users but not all; and
- **red** indicates compliance with none of the user requirements.

From those results, it can be concluded that the CRDP GODFIT v4 Level-2 total ozone column data records are compliant with most of the user requirements.

A proof-of-concept study of the validation of the ex-ante (prognostic) uncertainty estimates provided with the Level-2 TROPOMI data is presented in Section 7.

Table 4.2- Compliance of ERS-2 GOME GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	40 km along track
		320 km across track
Observation frequency	Daily – weekly	3 days at equator, 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	06/1995 – 06/2011
Total uncertainty	2% (radiative forcing studies)	Bias: 1.2%, Spread: ~3% (includes some co-location mismatch), Seasonality: 0.8%
	3% (variability studies)	
Dependences	–	SZA: 0.5-1.5 % up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 2% up to 240 K
Stability	1 – 3 % / decade	+0.2 % / decade



Table 4.3 - Compliance of Envisat SCIAMACHY GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	20 x 40 km ²
Observation frequency	Daily – weekly	3 days at equator, 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	08/2002 – 04/2012
Total uncertainty	2% (radiative forcing studies)	Bias: 0.7%, Spread: ~3% (includes some co-location mismatch) , Seasonality: 1.5-2 %
	3% (variability studies)	
Dependences	–	SZA: 0.5 - 1% up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 1% up to 240 K
Stability	1 – 3 % / decade	-0.6 to +0.4% / decade

Table 4.4 - Compliance of Aura OMI GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	13 x 24 km ²
Observation frequency	Daily – weekly	1 day at equator, < 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	10/2004 – 12/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 1.3%, Spread: ~2.7% (includes some co-location mismatch) , Seasonality: 0.9%
	3% (variability studies)	
Dependences	–	SZA: 1.5% up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 2% up to 240 K
Stability	1 – 3 % / decade	0% to +0.3%/decade

Table 4.5 - Compliance of MetOp-A GOME-2 GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	80 x 40 km ²
Observation frequency	Daily – weekly	1.5 days at equator, 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	01/2007 – 04/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 1%, Spread: ~2.8% (includes some co-location mismatch) , Seasonality: 1.2%
	3% (variability studies)	
Dependences	–	SZA: 1% up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 0.5% up to 240 K
Stability	1 – 3 % / decade	0.0% / decade



Table 4.6 - Compliance of OMPS SNPP GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	50 x 50 km ² at nadir
Observation frequency	Daily – weekly	Daily
Time period	(1980-2010) – (2003-2010)	1/2012 – 12/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 1.4%, Spread: 2.8% (includes some co-location mismatch) , Seasonality: 0.6%
	3% (variability studies)	
Dependences	–	SZA: 1.4% up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 2% up to 240 K
Stability	1 – 3 % / decade	-0.3 % / decade

Table 4.7 - Compliance of MetOp-B GOME-2 GODFIT v4 total ozone data with user requirements (URD 3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	80 x 40 km ²
Observation frequency	Daily – weekly	1.5 days at equator, 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	01/2013 – 11/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 1.5%, Spread: 2.7% (includes some co-location mismatch) , Seasonality: 1.0%
	3% (variability studies)	
Dependences	–	SZA: 1.3 % up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 3% up to 240 K
Stability	1 – 3 % / decade	1.0 %/decade

Table 4.8 - Compliance of TROPOMI/S5P OFFL total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	7 x 5 km ² 7 x 3.5 km ² , since August 2019
Observation frequency	Daily – weekly	Daily
Time period	(1980-2010) – (2003-2010)	11/2017 – 12/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 0.9%, Spread: ~2.8% (includes some co-location mismatch) , Seasonality: 1.5%
	3% (variability studies)	
Dependences	–	SZA: from 0.5 -1 % up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 2-2.5% up to 240 K
Stability	1 – 3 % / decade	N/A



Table 4.9 - Compliance of MetOp-C GOME-2 GODFIT v4 total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	80 x 40 km ²
Observation frequency	Daily – weekly	1.5 days at equator, 1 day at polar latitudes
Time period	(1980-2010) – (2003-2010)	02/2019 – 12/2021
Total uncertainty	2% (radiative forcing studies)	Bias: 1.3%, Spread: ~2.5% (includes some co-location mismatch) , Seasonality: 1.2%
	3% (variability studies)	
Dependences	–	SZA: ~1.3 % up to 70° SZA
		Latitude: negligible
		Clouds: no dependency on cloud cover
		Effective temperature: 2% up to 230 K
Stability	1 – 3 % / decade	N/A*

* The time period is too short for a statistically significant value to be extracted

4.3 Level-3 merged gridded total ozone

4.3.1 Level-3 GTO-ECV

One of the main aims of the ESA Ozone_cci+ project is to construct the homogeneous global long-term GOME-type Total Ozone Climate data record, hereafter termed GTO-ECV. The individual Level-2 observations (presented and validated above) are converted into a Level-3 product and then combined into one single cohesive record spanning the entire 27-years period from 1995 to 2021. This section summarizes the main characteristics of the merging methodology as well as the latest improvements and extensions implemented within the Ozone_cci+ project. A detailed description of the predecessor of GTO-ECV has been presented and validated in Loyola et al., 2009 [RD64] and Coldewey-Egbers et al., 2015 [RD34], whereas the current algorithm version and validation is extensively discussed in Garane et al., 2018 [RD38].

The individual Level-2 measurements processed with the GODFIT v4 retrieval algorithm are mapped onto a regular global grid of 1°x1° in latitude and longitude to construct daily averages for each sensor. Before combining the individual gridded data, adjustments are made in order to account for possible biases and drifts between the instruments. Figure 4.15, reproduced here from Garane et al., 2018 [RD38], shows the percentage differences between OMI and the other six sensors for 1° zonal monthly mean ozone columns during overlap periods. These zonal means were computed for co-located daily gridded data in order to minimize the impact of differences in the sampling pattern for OMI and the corresponding second sensor. In general, the inter-sensor consistency is very good; mean differences are between -0.2 ± 0.9 % (for GOME-2B, panel d) and 1.1 ± 1.3 % (for TROPOMI, panel e). In the inner tropics, the bias is slightly negative for all sensors and it increases toward higher latitudes. The differences between OMI and GOME show slightly larger scatter in the Southern Hemisphere due to significantly reduced spatial coverage of GOME because of the tape recorder failure in June 2003. The differences between OMI and SCIAMACHY indicate a positive bias for most parts of the Globe, with a maximum in the Southern Hemisphere around the polar night. For GOME, SCIAMACHY, TROPOMI/S5P, and GOME-2C we apply correction factors using the seasonal mean differences, calculated from the seasonal mean average of all available years, with respect to OMI as a function of latitude. The differences between OMI and GOME-2A indicate a positive drift of ~ 0.15 % per annum in the middle latitudes of both hemispheres, which we take into account during the adjustment. Likewise, for GOME-2B, the correction factors with respect to OMI depend on time (month) and latitude. The adjustment is then applied to the daily gridded data for each individual sensor. Thereby the monthly correction factors are linearly interpolated in time.



Subsequently, the individual data sets are combined into one single record. In contrast to the previous version (Coldewey-Egbers et al., 2015; [RD34]) where we used only one instrument at any given time, in GTO-ECV we now average all available daily measurements. GOME data are restricted to up and until December 2004. As the ground-based validation of SCIAMACHY Level-2 data indicates some lingering issues with the Level-2 TOCs we use SCIAMACHY only until December 2004 in order to fill the data gap between the GOME loss of global coverage and the launch date of OMI. GOME-2A is integrated until December 2017, due to the fact that the sensor started losing its solar visibility in 2018. The complete merged GTO-ECV data record with typical ozone characteristics is shown in Figure 4.16. Highest ozone values occur in northern hemispheric springtime, whereas monthly mean values are below 200 D.U. from September to November southwards of 70° S. The horizontal lines indicate the period for each sensor included.

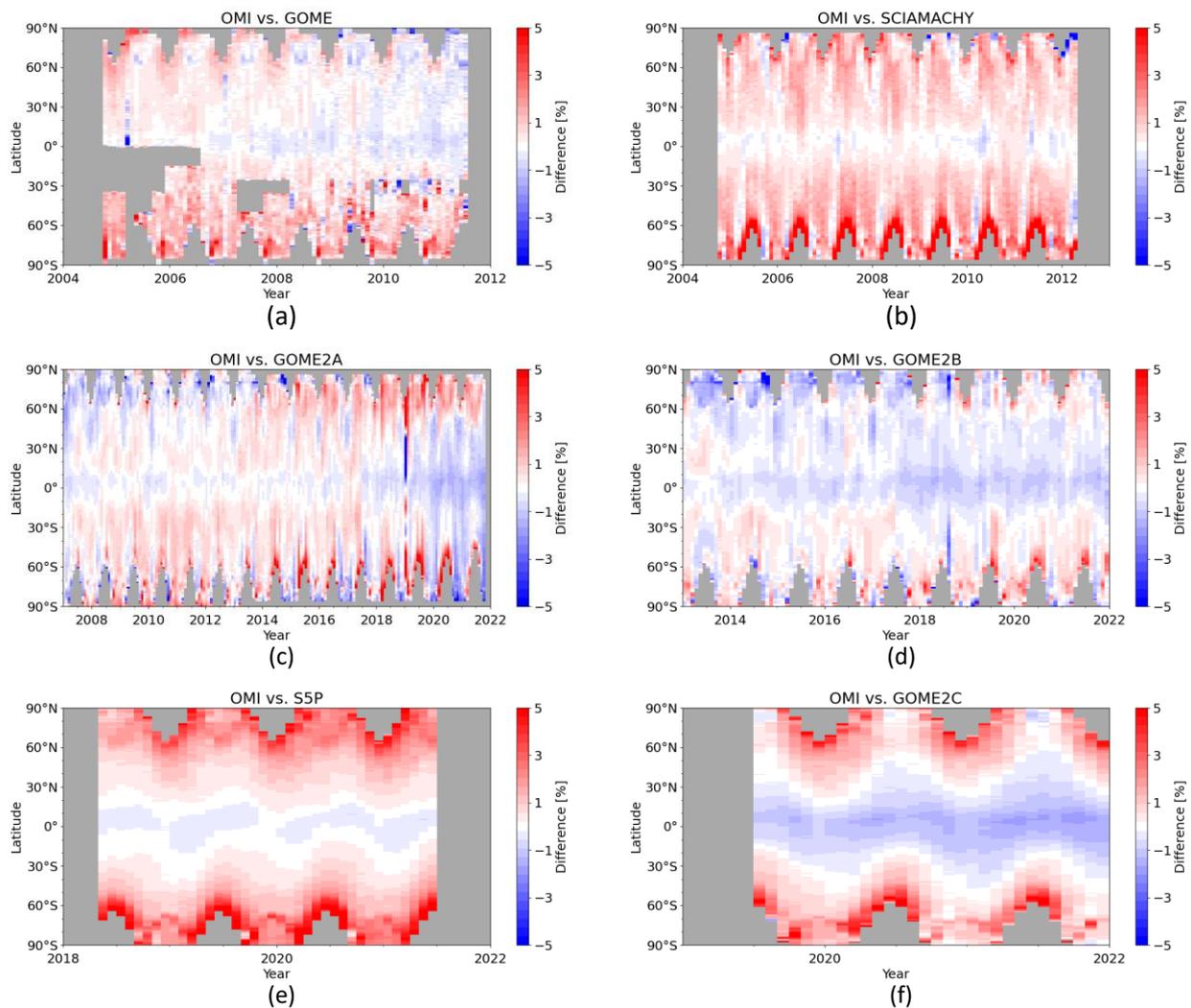


Figure 4.15 - Percentage differences between OMI and the other six sensors for 1° zonal monthly mean ozone columns during overlap periods. Panel (a): GOME; panel (b): SCIAMACHY; panel (c): GOME-2A; panel (d): GOME-2B; panel (e): TROPOMI/S5P and panel (f): GOME-2C.

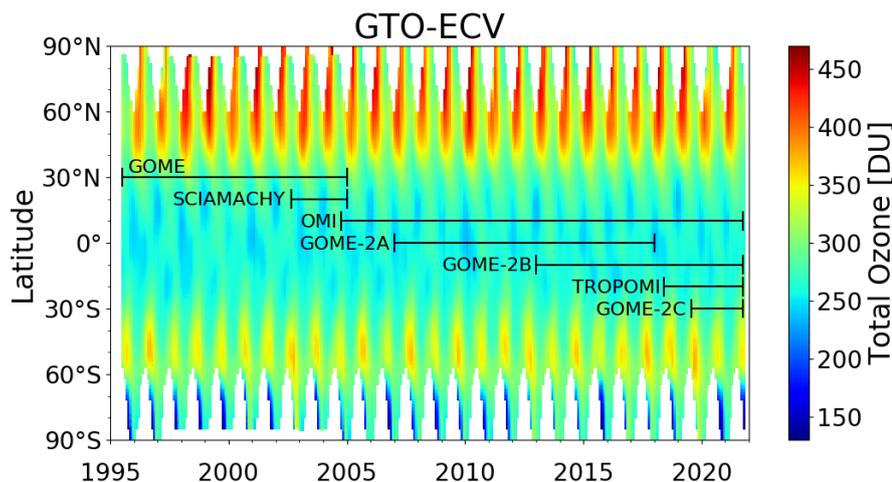


Figure 4.16 - GTO-ECV total ozone column data record as a function of latitude and time from 1995 to 2021. Black horizontal lines indicate the period for each sensor included in the merged product.

4.3.2 Level-3 validation results and discussion

The validation of the new Level-3 GTO-ECV merged product was performed using as ground truth the Brewer and Dobson spectrophotometer networks described in Section 3.5.1.1. In order to create the Level-3 TOC field, based on the WOUDC ground-based stations, the reported TOCs were gridded into the same $1^\circ \times 1^\circ$ grid as the GTO-ECV GODFIT data, on a monthly basis, with most grid points being represented by only one reporting station. In detail, direct Sun measurements were considered for the gridding of the ground-based TOCs into Level-3 grid points, even though in some cases this choice severely decreases the number of measurements. As a compromise between obtaining the highest global coverage possible and the most representative monthly means, especially at high latitudes, a lower limit of 10 measurements per month and per grid box was enforced so that the temporal representativeness errors are minimized.

Figure 4.17 shows the percentage difference between the satellite (Level-2 and Level-3) and the Dobson (left) and Brewer (right) TOC records, as a function of latitude. The seven individual satellite TOCs are very consistent with each other for all latitudes, within 2% for most latitude bins, and in very close agreement with the ground-based data. The Level-3 comparisons (blue diamonds) show very good agreement with the individual Level-2 latitudinal means. In particular, over the NH, all Level-2 show a positive deviation of 0 – 2 % to the ground-based data for both ground-based instrument types. TROPOMI/S5P and GOME-2C with respect to Brewer ground-based instruments show a different behaviour, especially for high Northern latitudes, but the different temporal coverage of the two sensors compared to the others should be taken under consideration. In the SH, the Level-3 comparisons are in very good agreement with the Level-2 comparisons, especially northwards 60° S. Below that latitude, the spread in comparisons reaches the 3.0 % level, which may be attributed to sampling differences between the Level-2 and Level-3 data (see Coldewey-Egbers et al., 2015 for more in-depth discussion of this issue).

In Figure 4.18, the NH and SH monthly mean time series comparisons of the Level-2 and Level-3 data records against the Dobson and Brewer measurements are shown. The Dobson comparisons for SH (panel c) and NH (panel a) show very good agreement between Level-3 and individual Level-2 lines, within the 1 % difference level for most of the 27-year data record, except for a small number of outliers. The agreement between the eight datasets and the ground-based measurements is excellent, with relative mean differences within 0.4 and 1.5%. For the entire time series of the Level-3 data record the mean relative difference remains mainly positive, with a mean relative bias of 0.8 – 1.4 %, depending in the hemisphere and the type of ground-based instrument. Concerning the Level-3 comparisons in the NH, the drift per decade of the differences with respect to ground-based data is negligible, 0.0 ± 0.1 % per decade for Dobson and -0.4 ± 0.1 % per decade



for Brewer co-locations. As it was shown in Section 4.2.6, the long-term drift in the differences of the individual Level-2 data sets utilized for the Level-3 data retrieval, was found to be always less than 1%/decade. The statistics for the Level-3 dataset are summarized in Table 4.10, last column.

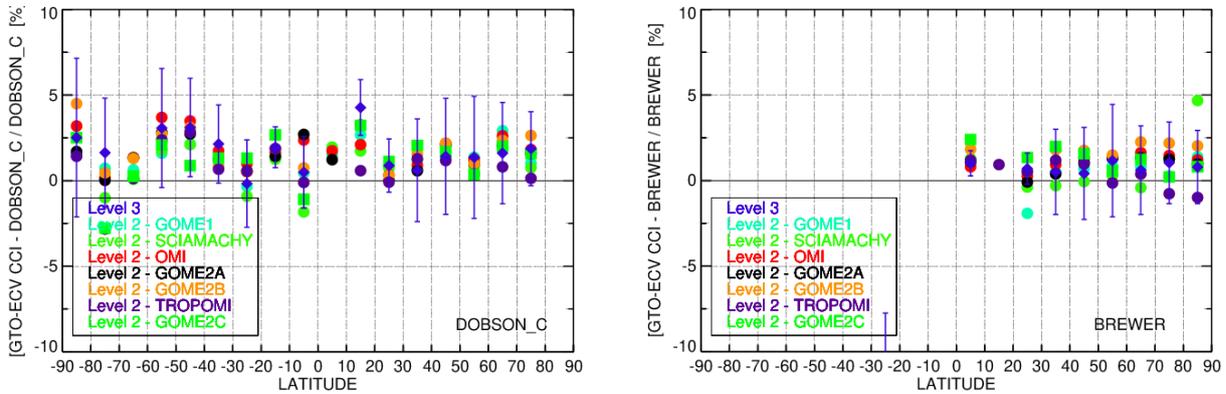


Figure 4.17 - Latitudinal variability of the percentage difference between satellite observations and ground-based measurements. Left: for the Dobson network and right: for the Brewer network. Cyan dots: GOME Level-2 comparison; green dots: SCIAMACHY Level-2 comparison; red dots: OMI Level-2 comparison; black dots: GOME-2A Level-2 comparison; orange dots: GOME-2B Level-2 comparison; purple dots: TROPOMI/S5P Level-2 comparison and green rectangles: GOME-2C comparison. Level-3 GTO-ECV comparison is shown with the blue diamonds. The 1- σ standard deviation of the average is also displayed only for the Level-3 lines.

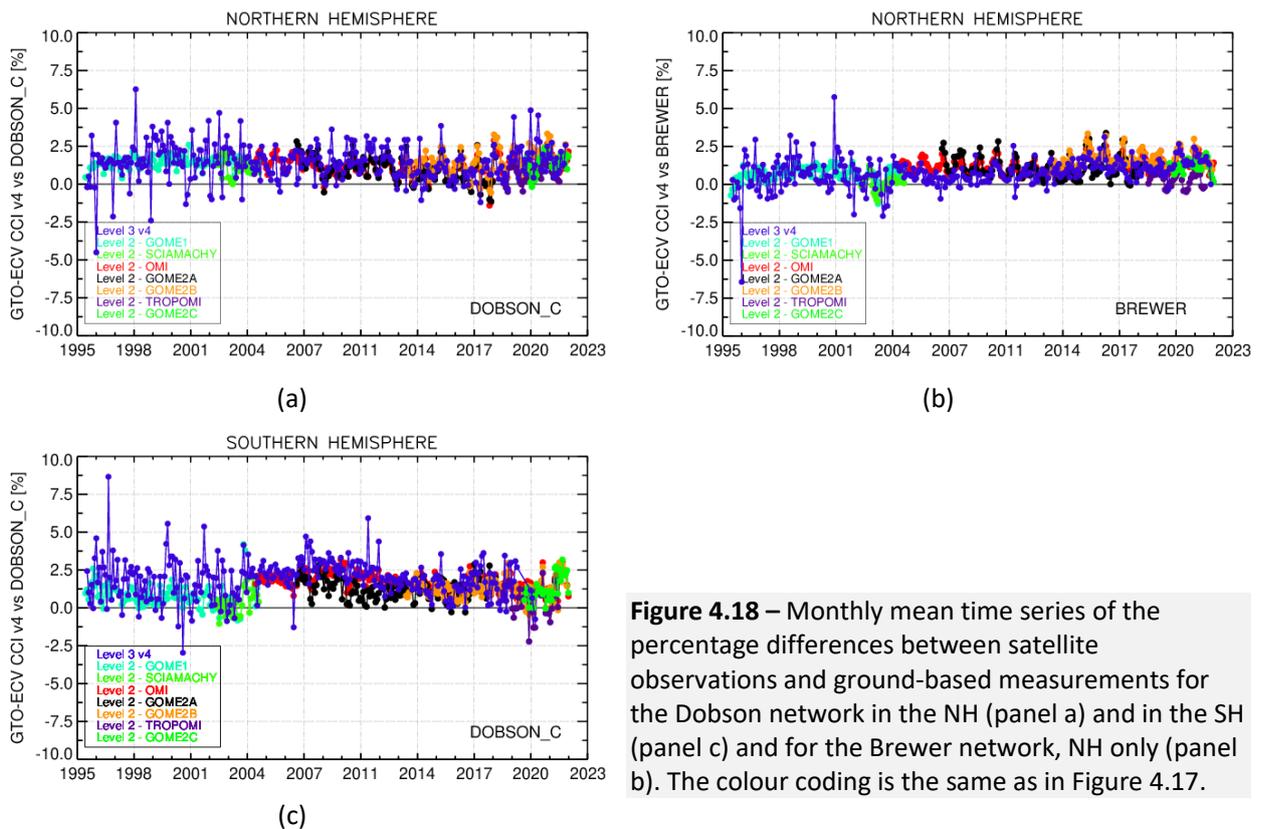


Figure 4.18 – Monthly mean time series of the percentage differences between satellite observations and ground-based measurements for the Dobson network in the NH (panel a) and in the SH (panel c) and for the Brewer network, NH only (panel b). The colour coding is the same as in Figure 4.17.

The good temporal stability of the GTO-ECV Level-3 TOC record, which well satisfies the requirements for the long term stability for total ozone measurements of between 1–3 % per decade (van der A et al., 2011) and the excellent inter-sensor consistency, make the new Level-3 GTO-ECV dataset suitable and useful for longer



term analysis of the ozone layer, such as decadal trend studies, the evaluation of chemistry-climate model projections, and data assimilation applications.

In order to assess and ensure the quality of the Level-3 GTO-ECV dataset, comparisons are also performed against the solar backscatter ultraviolet (SBUV) merged data product, which was quality assured in Frith et al., 2017, [RD37]. The continuity of the Level-2 GODFIT datasets and their agreement to the respective Level-2 SBUV/2 time-series was also shown in the work of Garane et al., 2018 [RD38]. In Figure 4.19, the monthly mean time series comparison between GTO-ECV and SBUV v8.7 merged total ozone product is presented for the NH and Dobson (panel a), the SH and Dobson (panel b) and the NH and Brewer (panel c) instrument types. The Level-3 GTO-ECV (red line) and SBUV merged (black line) datasets show a very good agreement of within $\pm 1.5\%$, considering their individual instrumental and algorithm differences, as well as a very similar seasonal variability over the entire time period, with a peak-to-peak amplitude $\sim 1\%$. The agreement is best for the NH, but for the SH co-locations the SBUV merged product agrees better to the Dobson ground-based measurements than the Level-3 GTO-ECV. Furthermore, the two datasets show an almost negligible drift per decade for both ground-based instrument networks:

- In the NH \rightarrow SBUV: -0.3% /decade, GODFIT Level-3: 0 to 0.4% /decade.
- In the SH \rightarrow SBUV: -0.4% /decade, GODFIT Level-3: 0.0% /decade.

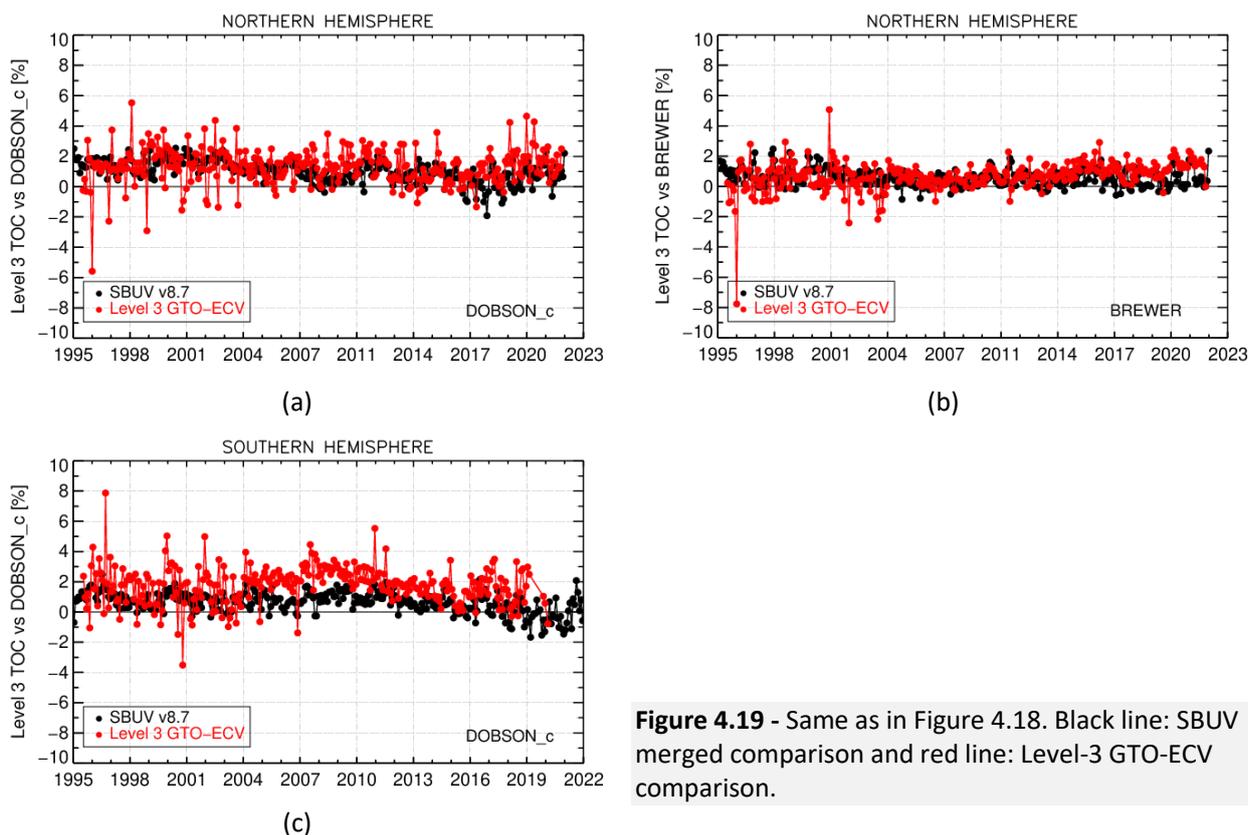


Figure 4.19 - Same as in Figure 4.18. Black line: SBUV merged comparison and red line: Level-3 GTO-ECV comparison.

4.3.3 Summary and compliance with user requirements

The GTO-ECV Level-3 data yield similar validation results as the ones obtained with the equivalent GODFIT v4 Level-2 validation of the individual GOME, SCIAMACHY, OMI, GOME-2A, GOME-2B, GOME-2C and TROPOMI/S5P Level-2 datasets. In detail:

- The Level-3 validation against ground-based measurements showed:
 - an excellent agreement to the ground-based measurements within 0.8 to 1.4% for the monthly mean time series,



- as well as a negligible drift in the Northern Hemisphere Brewer co-locations of -0.4 ± 0.1 %/decade.
- The individual Level-2 data sets show excellent inter-sensor consistency with mean differences within 1.0 % at moderate latitudes ($\pm 50^\circ$), whereas the Level-3 dataset shows mean differences (w.r.t. Level-2) that span between 0 and 1.3 %.

We hence conclude that the exceptional quality and temporal stability of the GTO-ECV Level-3 TOC record satisfies well the requirements of 1 – 3 % per decade. The inter-sensor consistency renders the Level-2 GOME/ERS-2, SCIAMACHY/Envisat, OMI/Aura, GOME2/Metop-A and GOME-2/Metop-B GODFIT v4 datasets, and the newly integrated sensors TROPOMI/S5P and GOME2/Metop-C, as well as the Level-3 GTO-ECV datasets, suitable and useful for longer term analysis of the ozone layer, such as decadal trend studies, the evaluation of model simulations, and data assimilation applications.



Table 4.10 - Overview of Level-2 & Level-3 validation results for the Ozone_cci+ CRDP total ozone column data products with respect to the Dobson (effective temperature corrected) and Brewer network.

		GOME/ ERS-2 (%)	SCIAMACHY / Envisat (%)	OMI/ Aura (%)	GOME-2/ Metop-A (%)	SNPP/ OMPS (%)	GOME-2/ Metop-B (%)	TROPOMI/ S5P (%)	GOME-2/ Metop-C (%)	Level-3 ECV v3	GTO-
Monthly mean bias and 1-sigma	Dobson*	1.5 ± 0.5	0.9 ± 0.7	1.3 ± 0.6	0.9 ± 0.7	1.4 ± 0.5	1.4 ± 0.6	1.1 ± 0.8	1.4 ± 0.7	1.4 ± 1.2	
	Brewer*	0.8 ± 0.5	0.4 ± 0.8	1.3 ± 0.4	1.2 ± 0.8	1.4 ± 0.4	1.7 ± 0.5	0.7 ± 0.7	1.2 ± 0.4	0.8 ± 0.9	
Monthly mean variability	Dobson*	±2.9	± 3.3	± 3.0	± 3.1	± 3.0	± 2.9	± 3.1	± 2.8	±3.1	
	Brewer*	± 2.7	± 2.7	± 2.4	± 2.5	± 2.5	± 2.4	± 2.5	± 2.2	±2.9	
Drift per decade	Dobson*	0.1 ± 0.1	-0.6 ± 0.1	0.0 ± 0.1	-0.1 ± 0.1	-0.1 ± 0.2	1.2 ± 0.2	0.9 ± 0.7	(1.7 ± 0.8)	0.0 ± 0.1	
	Brewer*	0.2 ± 0.1	0.4 ± 0.1	0.3 ± 0.1	0.0 ± 0.1	-0.5 ± 0.1	0.3 ± 0.2	0.9 ± 0.6	(1.9 ± 0.7)	-0.4 ± 0.1	
Seasonality (peak – to – peak)	Dobson*	0.8	1.5	0.9	1.0	0.6	0.9	1.7	1.4	1.1	
	Brewer*	0.8	2.0	0.9	1.4	0.7	1.0	1.3	0.9	0.9	
Latitude	Dobson	1.4 ± 0.8	1.1 ± 1.1	1.8 ± 1.0	1.5 ± 0.9	1.4 ± 1.0	1.8 ± 1.3	0.9 ± 1.4	1.2 ± 1.4	1.8 ± 1.1	
	Brewer*	0.7 ± 0.4	0.2 ± 0.5	1.2 ± 0.5	0.8 ± 0.5	1.4 ± 0.4	1.7 ± 0.6	0.5 ± 0.7	1.4 ± 0.7	0.8 ± 0.3	
Solar Zenith Angle	<70°	Dobson	1.3 ± 0.3	0.9 ± 0.5	1.8 ± 0.7	1.1 ± 0.6	1.5 ± 0.4	1.2 ± 0.7	0.5 ± 1.2	1.3 ± 0.2	N/A
		Brewer*	0.7 ± 0.3	0.0 ± 0.7	1.2 ± 0.4	0.7 ± 0.6	1.3 ± 0.3	1.5 ± 0.5	0.9 ± 0.6	1.3 ± 0.4	N/A
	>70°	Dobson	1.2 ± 1.2	-0.8 ± 3.0	1.0 ± 2.0	1.6 ± 0.2	-0.1 ± 1.0	1.1 ± 2.0	0.0 ± 2.0	-0.7 ± 2.6	N/A
		Brewer*	0.2 ± 1.5	-0.4 ± 2.5	2.2 ± 0.1	-1.8 ± 6.3	0.5 ± 0.8	2.5 ± 0.2	-1.2 ± 1.5	0.0 ± 0.8	N/A

* Only Northern Hemisphere.



4.4 Level-2 total ozone from IASI-A, -B and -C

In addition to the Level-2 total ozone records that are retrieved by the GODFIT v4 algorithm, total ozone observations performed by the IASI instruments on board the MetopA, MetopB and MetopC satellites (hereafter IASI-A, IASI-B and IASI-C), are also validated against ground-based measurements. The current IASI total ozone retrieval algorithm is the FORLI-O3 (Fast Optimal/Operational Retrieval on Layers for IASI) v20191122 [RD4], which is applied since December 2019. The previous version, FORLI-O3 v20151001, is used for total ozone data retrievals before that time and was validated by Boynard et al. (2018) [RD30]. The back-processing of the dataset with the new version of the algorithm is currently ongoing. The IASI total ozone data are available via the AERIS portal (<https://iasi.aeris-data.fr/O3/>), but for the purposes of this validation work the ground-based stations overpass files with maximum co-location search radius up to 50 km, were provided directly by ULB/LATMOS. Table 4.11 shows the time span of the available Level-2 datasets.

To ensure the good quality for the observations, the provided data were already filtered by the following criteria:

- Cloud fraction: 0-13% (only pixels with a cloud fraction equal to or lower than 13 % are processed)
- Degrees of Freedom (DOF): 2-5, to exclude bad quality data mainly from the in the Antarctic region (DOF<2).
- The spectral fit residual root mean square error (RMS) is always less than $3.5 \times 10^{-8} \text{ W/m}^2/\text{cm}^{-1}$, excluding the cases where the difference between observed and simulated radiances is too high.

Additionally, to limit the noise in the validation results the O₃ integrated relative error was restricted to values equal or lower than 2 %. As a result, almost 5 % of the co-locations to ground-based total ozone measurements were excluded, mainly originating from the Antarctic. Figure 4.20 shows the latitudinal dependence of the relative percentage differences between the satellite and Dobson ground-based observations, averaged in 10° latitude bins. The comparisons of all three sensors southwards 80°S with extremely high percentage differences, spanning 25–35%, are disregarded when the ozone integrated relative error is limited to 2%.

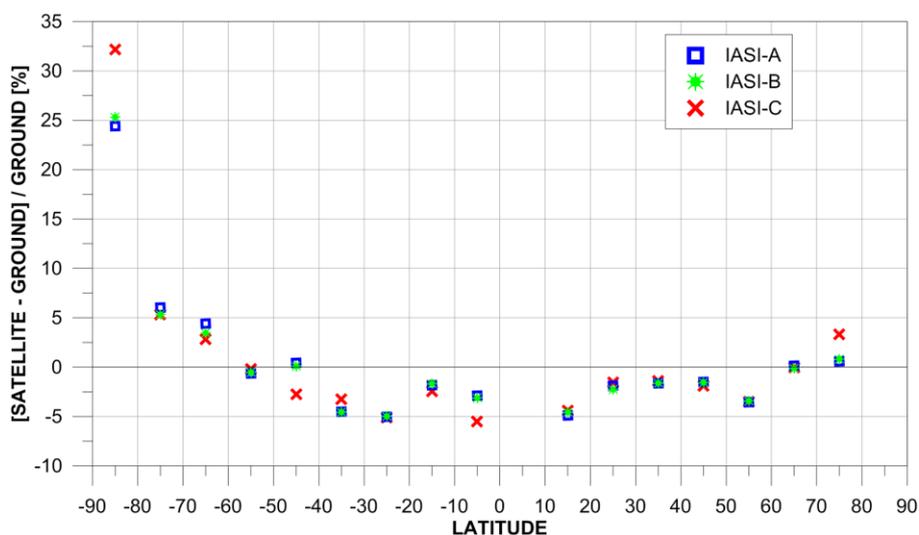


Figure 4.20 - The latitudinal dependence of the relative percentage differences between the IASI-A (blue symbols), IASI-B (green symbols) and IASI-C (red symbols) and Dobson ground-based observations, averaged in 10° latitude bins.



Table 4.11 - The time span of the available Level-2 IASI datasets.

Sensor	Algorithm	Time span	
IASI – MetopA*	Forli v20191122	5/2015 – 12/2020	5.5 years
IASI - MetopB		5/2015 – 12/2020	5.5 years
IASI - MetopC		9/2019 - 12/2020	16 months

*6 months missing (Jan-June 2019)

In the following sections, the relative percentage differences between the co-located satellite and ground-based total ozone observations will be investigated in terms of systematic bias. The temporal and geographical variation of the bias will be studied, as well as its dependence on various influence quantities. In parallel, the inter-sensor consistency of the three IASI instruments will be studied.

4.4.1 Systematic bias and its variations

In Figure 4.21, the histograms (left column of plots) and scatter plots (right column) of the comparisons between the three IASI instruments and the co-located Brewer ground-based total ozone measurements, representing the NH only, are shown. IASI-A comparisons are shown in panels a and b, IASI-B in panels c and d, and IASI-C in panels e and f. The histograms of the three sensors show normal distributions of the relative percentage differences around the mean bias, which ranges between -1.6 and -1.8 % for the Brewer co-locations. The respective Dobson comparisons, which cover both hemispheres, are very similar and are not shown here. Nevertheless, the statistics of the analysis with respect to both Brewer and Dobson ground-based instruments are summarized in Table 4.12. The mean relative bias with respect to Dobsons is also negative for the three IASI instruments, ranging between -2.3 and -2.5 %. The higher bias comes from the SH co-locations of the Dobson observations, as will be shown in the following. The main result thus far is that the three IASI sensors report lower TOCs than the ground-based networks by ~-2.5 to -1.5%. The scatter plots show that the correlation coefficient between satellite measurements and the Dobsons is 0.91 - 0.92, while the co-locations to Brewer observations are in even better agreement with the IASI sensors, with a correlation coefficient of 0.94 – 0.95.

To examine the temporal stability of the three satellite sensors' validation results, the hemispheric monthly mean time series of their relative percentage differences, seen in Figure 4.22, were exploited, where the timeseries of the comparisons between IASI-A and ground-based observations are shown with the blue line and symbols, IASI-B is shown with the green line and symbols, and IASI-C with red line and symbols. Panels a and c show the comparisons to Dobson measurements (panel a for the NH and panel c for the SH), while panel b shows the Brewer comparisons for the NH only.

Table 4.12 - The overall statistics that result from the co-locations of the three satellite sensors to the Brewer (NH only) and Dobson ground-based observations.

		IASI-A	IASI-B	IASI-C
Mean bias (%)	Dobson	-2.3	-2.3	-2.5
	Brewer	-1.6	-1.7	-1.8
St. Deviation (%)	Dobson	5.5	5.5	5.6
	Brewer	4.7	4.6	4.6
R ²	Dobson	0.92	0.92	0.91
	Brewer	0.95	0.95	0.94
N. of co-locations	Dobson	27981	29728	5879
	Brewer	45112	52010	11610

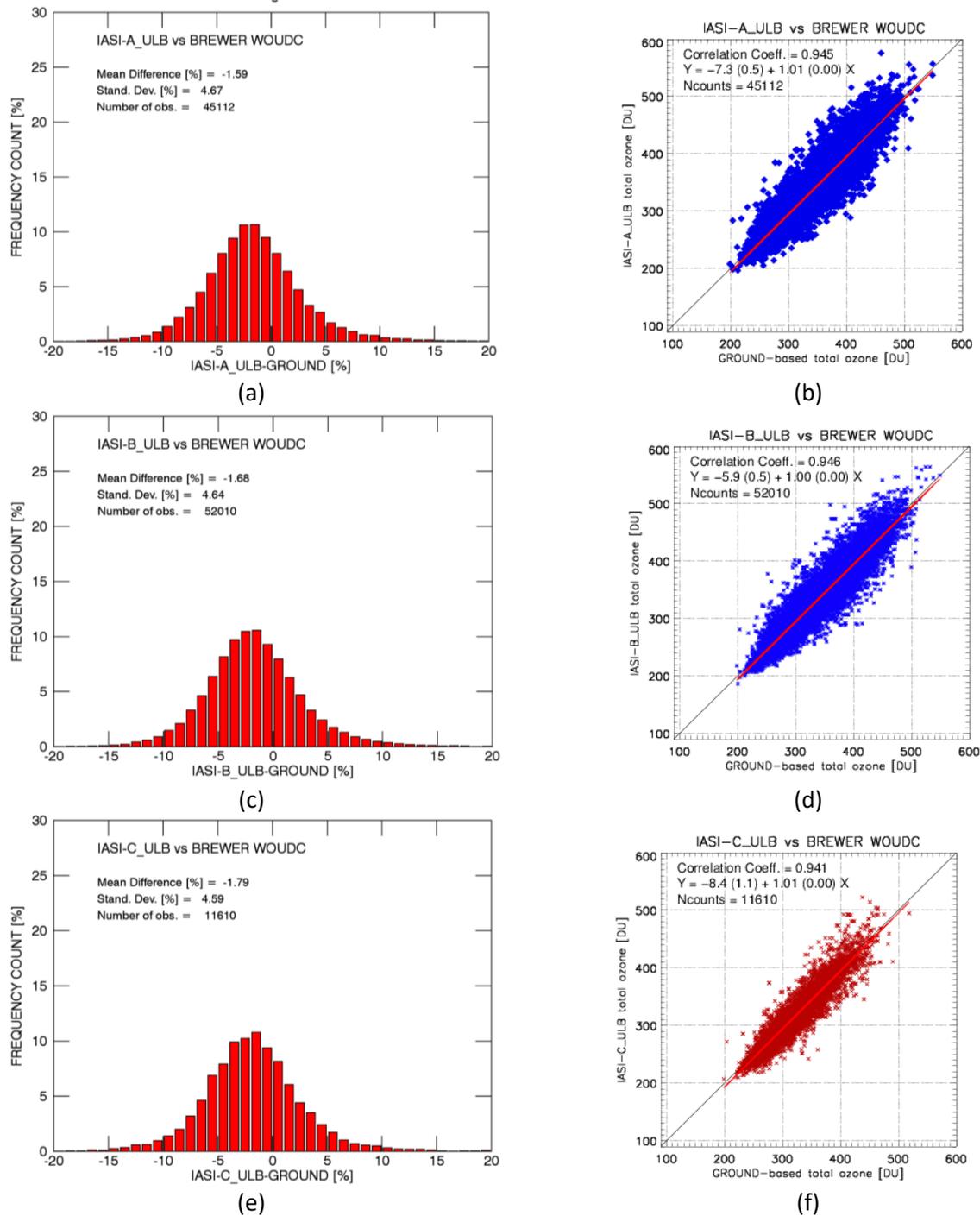


Figure 4.21 - The histograms (left column of plots) and scatter plots (right column) of the comparisons between the three IASI (IASI-A in panels a & b; IASI-B in panels c & d; IASI-C in panels e & f) instruments and the co-located Brewer ground-based total ozone measurements, representing the NH only.

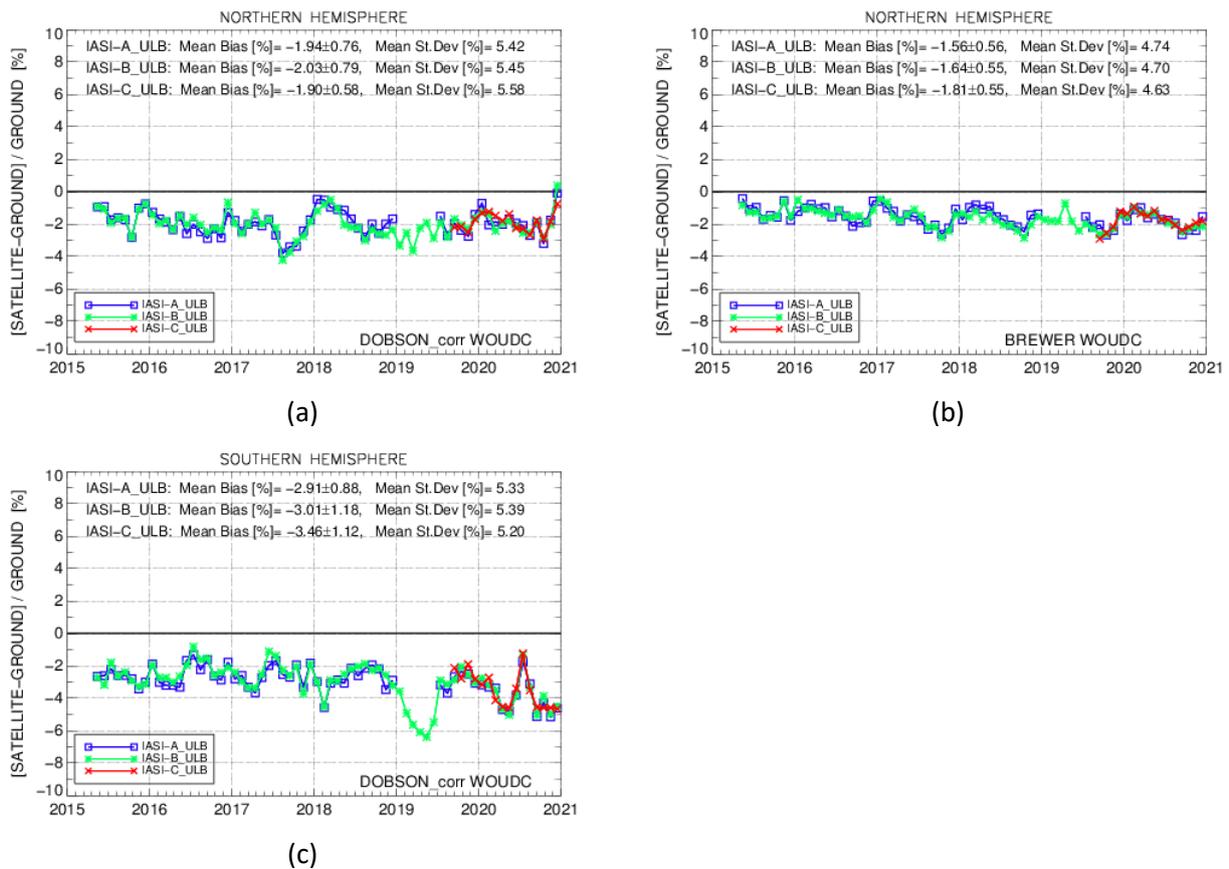


Figure 4.22 - The hemispheric monthly mean time series of the relative percentage differences between IASI-A (blue line and symbols), IASI-B (green line and symbols) and IASI-C (red line and symbols) and ground-based observations (panels a & c: Dobson, NH and SH, respectively; panel b: Brewer, NH).

The temporal consistency between the three IASI sensors for the time period of their operation in tandem¹ is remarkable for both hemispheres. The NH co-locations for the time period May 2015 – December 2020, for both types of ground-based instruments (panels a and b), are temporally very stable and follow a very similar pattern with lower differences, about 0 to -1%, during winter and spring months and higher during summer, up to -2 to -3 % with respect to ground-truth. The seasonal dependency of the relative percentage differences is also shown in Figure 4.23, with the respective peak-to-peak percentage ranges per sensor and type of ground-based instrument, summarized in Table 4.13.

The overall hemispheric mean relative bias is ranging between -2.0 ± 0.8 % and -1.6 ± 0.6 %, and the mean standard deviation of the monthly mean relative differences is ~ 4.7 % for the Brewer and ~ 5.4 % for the Dobson comparisons. In the SH, where only Dobson ground-based observations are available, the time series is also very stable temporally, showing a similar seasonal pattern (better seen in Figure 4.23, panel c) with higher differences during local summer months up to -3.5 % and lower during local winter and spring, up to -2 %. The only exception is the first semester of 2019, for which only IASI-B data are available and they show higher discrepancies in the SH, namely up to -6.3 % for May.

In Figure 4.24 the percentage differences of the co-locations are averaged in 10° latitude bins using each station's latitude as reference and plotted as a pole-to-pole graph for the Dobson (left panel) and the Brewer (right panel) stations. A latitudinal dependency of the co-locations is seen for both ground-based networks, showing a negative relative bias of -2 to -4 % for the tropics and low mid-latitudes. For mid and high latitudes

¹ The IASI-A total ozone column dataset available for validation was missing the first six months of data for 2019.



in the NH the relative biases are also negative but lower ($\sim -0.5\%$ for the Brewer co-locations) or even positive of up to $+1\%$ (for the Dobson co-locations). As explained above, the O_3 integrated relative error filter ($\leq 2\%$) that was applied to the datasets excluded almost all co-locations southwards $80^\circ S$, which had very high positive relative biases, of $\sim 25-35\%$. It should also be noted that the temporal coverage of the IASI-C co-locations is different than the other two sensors, resulting to divergencies in some latitude belts, e.g. 0 to $10^\circ S$.

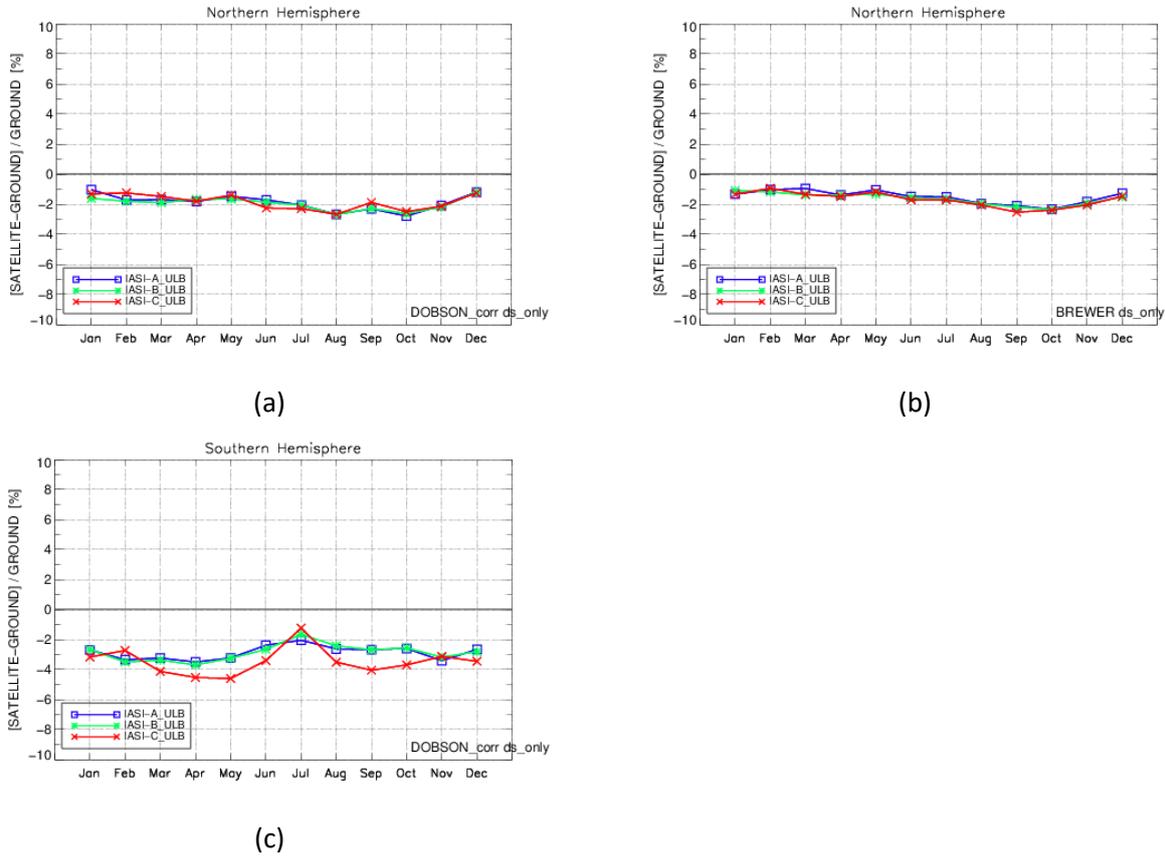


Figure 4.23 – The seasonal dependence of the three IASI sensors with respect to co-located ground-based total ozone measurements (panels and colour codes as in Figure 4.22).

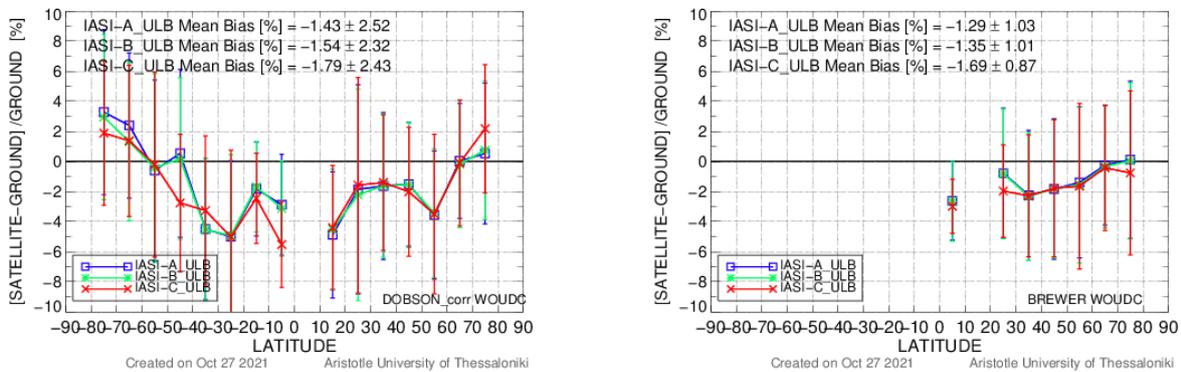


Figure 4.24 - The latitudinal dependency of the percentage differences between the three IASI sensors (colour coding as in Figure 4.22) and ground-based observations (left panel: Dobson; right panel: Brewer), averaged in 10° latitude bins.



4.4.2 Dependency on various parameters of the retrieval algorithm

The influence of various parameters that affect the satellite total ozone retrievals is also investigated, by plotting the percentage differences of the co-locations with respect to the parameter in question (Figure 4.25). Only comparisons to Brewer total ozone observations will be shown, because their number of co-locations (Table 4.12) to the IASI sensors is higher than those against the Dobson network. Moreover, it was seen that the Dobson results do not differ significantly from those shown here. In Figure 4.25, some data points are assigned numbers that appear at the top of the plot. This means that for those particular data points, the number of co-locations that correspond to the particular averaging bin is less than 5% of the total, indicating the significance of the dependency, if any.

Solar Zenith Angle (SZA)

The dependency of the percentage differences of the three IASI sensors and Brewer total ozone measurements on SZA is shown in panel a. The mean differences are always negative for all SZAs, ranging from -1% for $20^\circ < \text{SZAs} < 30^\circ$ and $\text{SZAs} > 70^\circ$, to -2 % for moderate SZAs (40° - 60°). This conclusion is applicable to all satellite sensors.

Pixel of the scan

Panel b shows that there is no dependence of the satellite and ground-based measurements comparisons on the satellite pixel.

Surface temperature

The dependence of the percentage differences on the surface temperature (in K), which is an input parameter to the Radiative Transfer Model used for the retrievals, is shown in panel c. For temperatures above 260 K, there is a U-shaped dependence of the comparisons on surface temperature, ranging between -3 % and 1.5 %. When the temperatures become extremely low, below 250 K, the percentage differences obviously representing the co-locations from high latitude stations, become positive up to +9 %.

Ozone profiles Degrees of Freedom (DOF)

Panel d shows the dependence of the percentage differences on the Degrees of Freedom of the signal, which is a quality flag for the data under investigation. As mentioned above, only data with $\text{DOF} > 2$ are processed to avoid bad quality observations from the Antarctic area. The dependence of the co-locations on the DOF shows that for values below 2.5 the corresponding datapoints result from a very low number of co-locations but introduce high differences, up to +6 %. We suggest that these data could be excluded to limit the noise in the observations.

Cloud cover in the pixel

Finally, panel e shows the total ozone retrievals from the three sensors do not depend on cloud cover (which is a filtering criterion, leaving only clear skies observations within the dataset), since no variability is observed for the averaged bins, that correspond to very low cloud coverage in the field of view of the measurement.

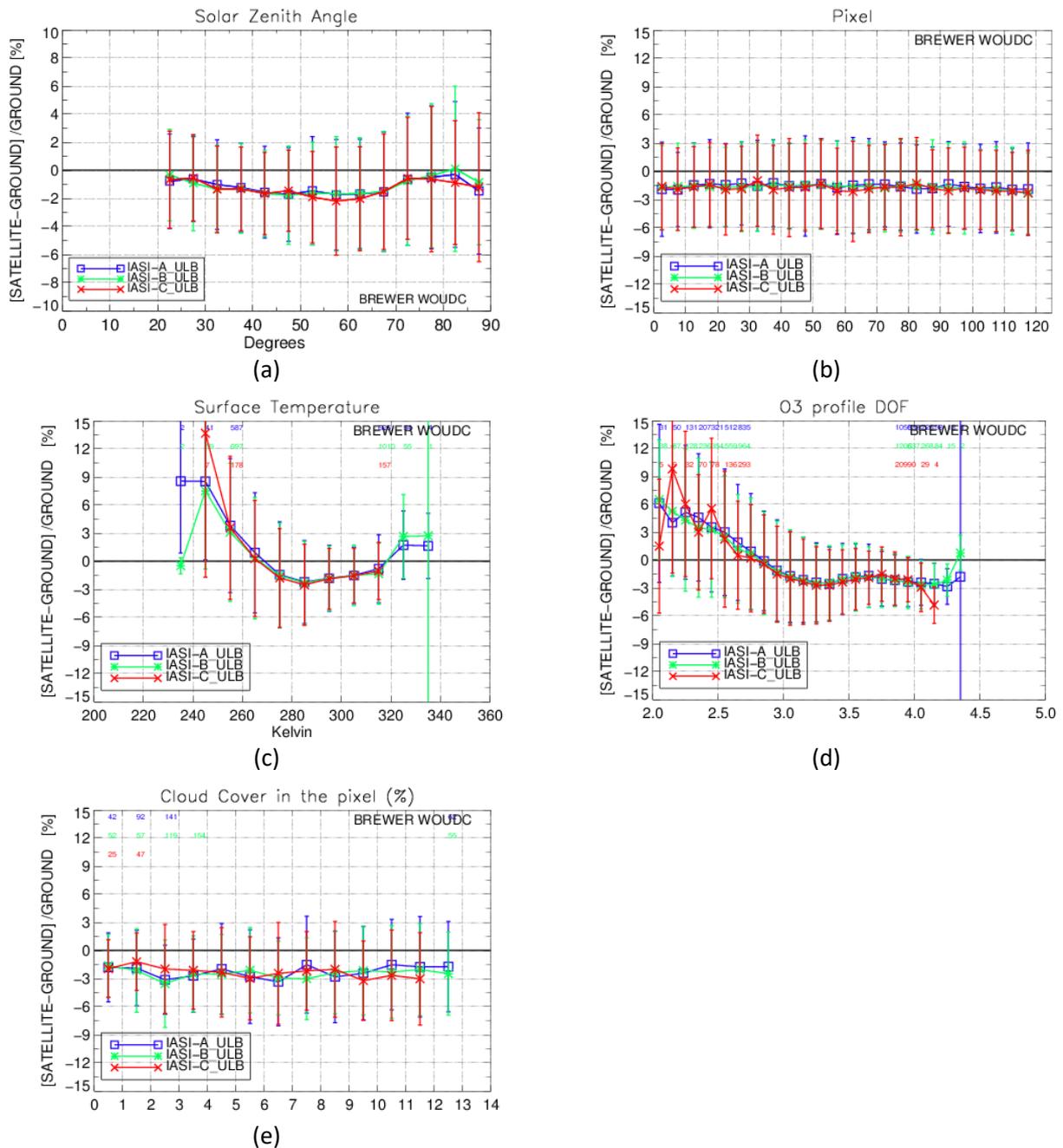


Figure 4.25 – The dependence of the relative percentage differences between the three IASI sensors (colour coding as in Figure 4.22) and Brewer ground-based total ozone observations, on various influence parameters, such as solar zenith angle (panel a), pixel number of the sensor (panel b), surface temperature (panel c), the number of degrees of freedom (panel d) and cloud coverage (pixel e).



4.4.3 Summary and compliance with user requirements

Total ozone columns retrieved from IASI-A, IASI-B and IASI-C with the FORLI-O3 v20191122 algorithm, were validated against Brewer and Dobson ground-based measurements. The time span of the IASI observations is 5.5 years for IASI-A and IASI-B, and only 16 months for IASI-C. The validation results, that are displayed in Table 4.13, can be summarized to the following points:

- the individual Level-2 IASI data sets show exceptional inter-sensor consistency with mean differences less than 0.1%;
- the mean relative bias between satellite and Brewer and Dobson reported TOCs is well within -1.5 and -2.5 % for all sensors, showing that the IASI sensors report lower TOCs than the ground-based measurements of both networks;
- The peak-to-peak seasonality of the relative differences ranges between 1.3 and 1.8 %.

The requirements that have to be met by the retrieved total ozone columns from the three IASI sensors are determined by the international climate-chemistry modelling community and listed in [RD8]. Namely, the stability of the TOC measurements has to be between 1 and 3% per decade, the radiative forcing introduced by the evolution of the ozone layer has to be less than 2% and that the short-term variability has to be less than 3 %. Due to the limited length of the available dataset, the stability requirement cannot be studied. As in the previous sections, the key validation results for the IASI-A, IASI-B and IASI-C data records are summarised in Table 4.14 to Table 4.16, respectively. These tables reproduce data quality criteria established by the Climate Research Group (CRG) in the Ozone_cci+ User Requirement Document [RD8]. The level of compliance of the IASI datasets with these user requirements is highlighted with the same a colour code as before:

- **green** indicates ascertained compliance with requirements from all contributing users;
- **yellow** indicates compliance with requirements from some users but not all; and
- **red** indicates compliance with none of the user requirements.

From those results, it can be concluded that the IASI Level-2 total ozone column data records are compliant with the requirements by most users.



Table 4.13: Overview of Level-2 validation results for the IASI total ozone column data products with respect to the Dobson (effective temperature corrected) and Brewer network.

		IASI-A (%)	IASI-B (%)	IASI-C (%)	
Monthly mean bias and 1-sigma	Dobson*	-1.9 ± 0.8	-2.0 ± 0.8	-1.9 ± 0.6	
	Brewer*	-1.6 ± 0.6	-1.6 ± 0.6	-1.8 ± 0.6	
Monthly mean variability	Dobson*	5.4	5.5	5.6	
	Brewer*	4.7	4.7	4.6	
Seasonality (peak – to – peak)	Dobson*	1.8	1.5	1.4	
	Brewer*	1.4	1.3	1.6	
Latitude	Dobson	-1.4 ± 2.5	-1.5 ± 2.3	-1.8 ± 2.4	
	Brewer*	-1.3 ± 1.0	-1.4 ± 1.0	-1.7 ± 0.9	
Solar Zenith Angle	<70°	Dobson	-2.1 ± 0.4	-2.0 ± 0.5	-2.2 ± 0.6
		Brewer*	-1.3 ± 0.4	-1.4 ± 0.4	-1.5 ± 0.5
	>70°	Dobson	-0.2 ± 0.9	-0.1 ± 0.6	0.0 ± 0.6
		Brewer*	-0.7 ± 0.5	-0.5 ± 0.4	-0.8 ± 0.3

* Only Northern Hemisphere.

Table 4.14 - Compliance of IASI MetopA FORLI-O3 v20191122 total ozone data with user requirements (URD v3.1)

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	50 km along track
		50 km across track
Observation frequency	Daily – weekly	Twice a day
Time period	(1980-2010) – (2003-2010)	05/2015 – 12/2020
Total uncertainty	2% (radiative forcing studies)	Bias: -2.0%, Spread: ~5% (includes some co-location mismatch), Seasonality: 1.5%
	3% (variability studies)	
Dependences	–	SZA: -2.0 % up to 70° SZA
		Latitude: significant (~6% peak-to-peak between tropics and high latitudes)
		Clouds: no dependency on cloud cover
		Surface temperature: 6% peak-to-peak above 250 K



Table 4.15 - Compliance of IASI MetopB FORLI-O3 v20191122 total ozone data with user requirements (URD v3.1)

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	50 km along track
		50 km across track
Observation frequency	Daily – weekly	Twice a day
Time period	(1980-2010) – (2003-2010)	05/2015 – 12/2020
Total uncertainty	2% (radiative forcing studies)	Bias: -2.0%, Spread: ~5% (includes some co-location mismatch), Seasonality: 1.4%
	3% (variability studies)	
Dependences	–	SZA: -2.0 % up to 70° SZA
		Latitude: significant (~6% peak-to-peak between tropics and high latitudes)
		Clouds: no dependency on cloud cover
		Surface temperature: 6% peak-to-peak above 250 K

Table 4.16 - Compliance of IASI MetopC FORLI-O3 v20191122 total ozone data with user requirements (URD v3.1)

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	50 km along track
		50 km across track
Observation frequency	Daily – weekly	Twice a day
Time period	(1980-2010) – (2003-2010)	09/2019 – 12/2020
Total uncertainty	2% (radiative forcing studies)	Bias: -1.9%, Spread: ~5% (includes some co-location mismatch), Seasonality: 1.5%
	3% (variability studies)	
Dependences	–	SZA: -2.0 % up to 70° SZA
		Latitude: significant (~6% peak-to-peak between tropics and high latitudes)
		Clouds: no dependency on cloud cover
		Surface temperature: 6% peak-to-peak above 250 K

4.5 Level-4 assimilated total ozone

4.5.1 The MSR Level-4 total ozone product

The MSR product is a multi-decadal Level-4 (i.e. assimilation-based) total ozone column data record based on most available ozone column satellite data sets, ground-based Brewer and Dobson observations, and an assimilation scheme with detailed error modelling. It is produced in two steps: First, the available ozone column satellite data sets are corrected for biases as a function of solar zenith angle (SZA), viewing zenith angle (VZA), time (trend), and stratospheric temperature using ground-based observations of the ozone column from Brewer and Dobson spectrophotometers from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC). Subsequently the de-biased satellite observations are assimilated within the ozone chemistry and data assimilation model TMDAM. For the data set validated here, the latest total ozone



retrievals of 15 satellite instruments are used: BUV-Nimbus4, TOMS-Nimbus7, TOMS-EP, SBUV-7, -9, -11, -14, -16, -17, -18, -19, GOME, SCIAMACHY, OMI and GOME-2. The time coverage is being extended backward in time using solely ground-based data, but the currently validated data sets starts in 1979. It is also still based on meteorology from the ERA-Interim reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), but an upgrade to ERA-5 is under construction. Temporal and spatial resolution of the current data set (from 1979 onwards) are 1 month and 0.5 by 0.5 deg² respectively.

4.5.2 Level-4 validation results and discussion

As the MSR product uses ground-based Brewer and Dobson measurements in the bias-correction procedure, these cannot be considered independent reference data. The validation will therefore be done solely with the NDACC ZSL-DOAS instrument network. The global distribution of the instruments yielding a sufficient amount of co-locations is shown in Figure 4.26, where sufficient is defined as follows:

- at least 10 years spanned by the co-locations,
- at least 10 ZSL-DOAS measurements contribute to a ground-based monthly mean at a station, and
- the effective day differs less than 5 days from the middle of the month (usually only problematic at the beginning and end of polar day/night).

Co-locations are defined as: same month and the MSR grid cell contains the station. No stations are averaged, meaning that even if multiple stations fall within the same MSR grid cell, they are treated separately.

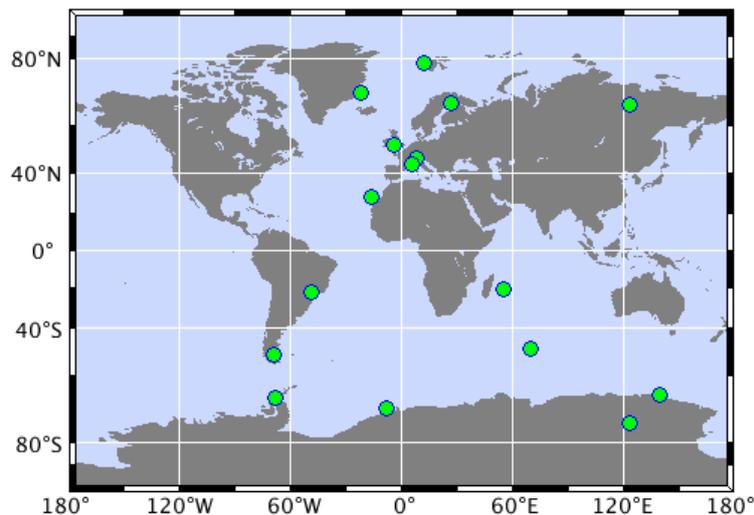


Figure 4.26 - Global distribution of the ZSL-DOAS instruments used to validate the MSR Level-4 total ozone product.

As an example, the time series of MSR and co-located ZSL-DOAS total columns of ozone at the Observatoire de Haute Provence, France, are shown in Figure 4.27, and the corresponding analysis of the differences, in terms of median, spread, and drift, in Figure 4.28.

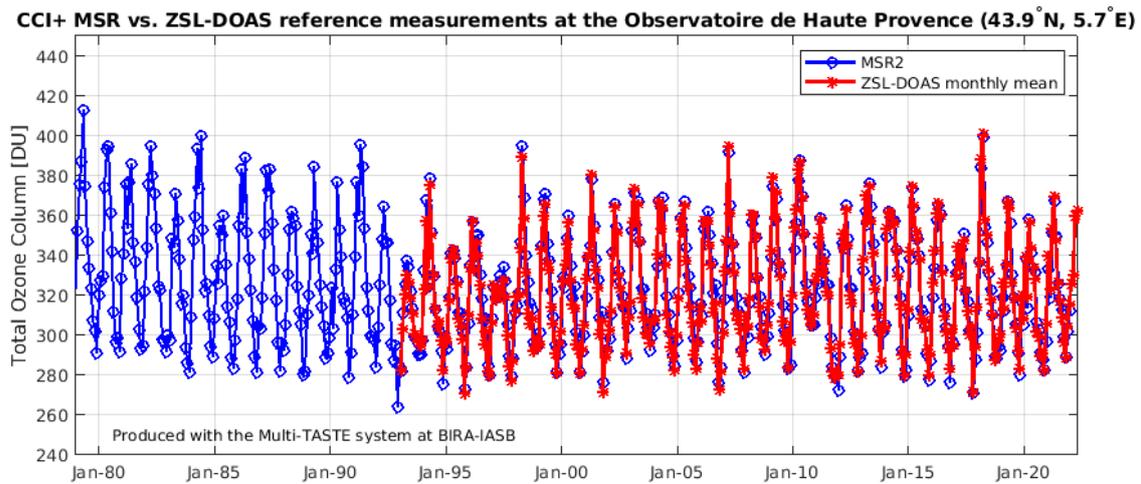


Figure 4.27 - Time series of monthly mean total ozone, from the Level-4 MSR product (blue) and from co-located SAOZ measurements at the Observatoire de Haute Provence, France (red).

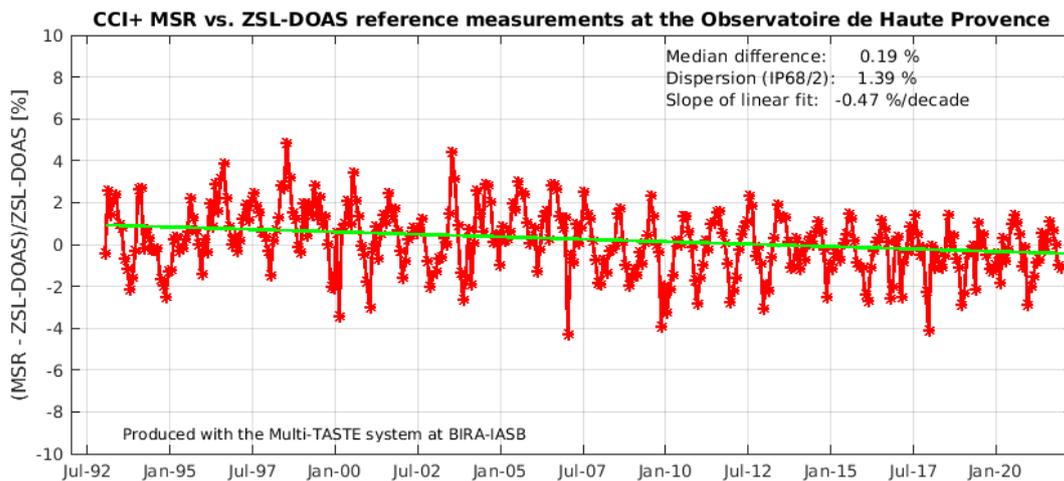


Figure 4.28 - Differences in monthly mean total ozone at OHP, corresponding to the time series in Figure 4.27. Key statistics are also provided.

This analysis is performed for all 16 instruments yielding a sufficient amount of co-location and the results are presented as a function of instrument latitude in Figure 4.29. On the global scale (i.e. averaging the statistics over the different stations), the MSR data set easily satisfies the user requirements. Looking at individual stations, user requirements are in general also met, but at high latitudes both the mean difference and the comparison spread show larger deviations from zero. It must be kept in mind that these comparisons do not yet involve the use of 2-D observation operators to improve the spatial co-location, and some co-location mismatch uncertainty is therefore affecting these comparisons, especially in polar regions where strong gradients in the ozone field are often present (due to polar vortices for instance). This will be improved upon for the next version of the PVIR.

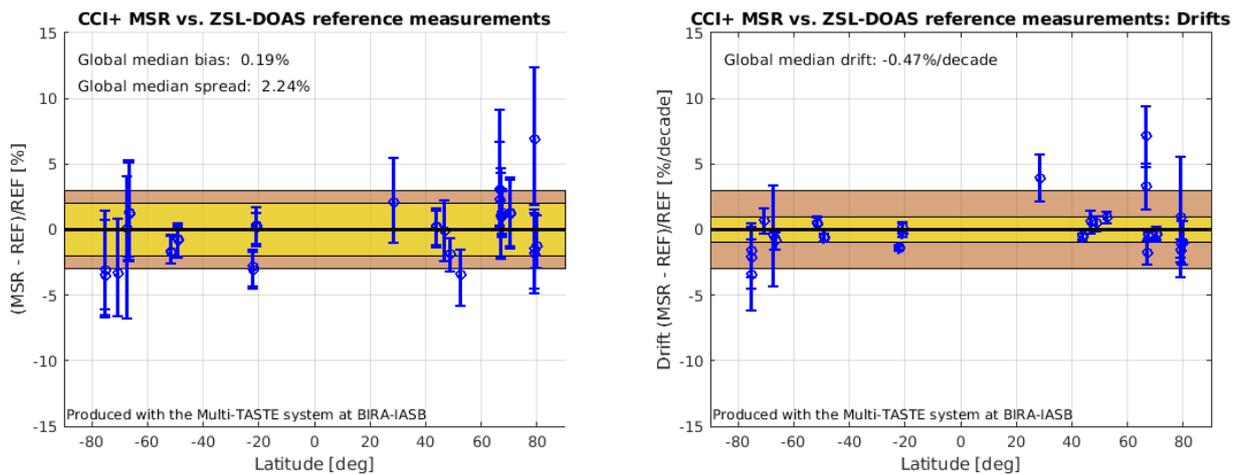


Figure 4.29 - Pole-to-pole visualisation of (1) the bias and spread per ZSL-DOAS instrument in the left-hand panel, and (2) the long-term drift in the right-hand panel. User requirements are indicated by the coloured regions (i.e. a maximum of 2-3% total uncertainty and a maximum of 1-3%/decade drift).

4.5.3 Summary and compliance with user requirements

The MSR Level-4 assimilated total ozone column product was validated with fully independent ZSL-DOAS measurements obtained from 16 different instruments spread across the globe, with the earliest measurements made in the late eighties (leaving the early-eighties MSR data unvalidated). The median bias over the network of 0.19% and the median comparison spread of 2.24% are well within the user requirements, and so is the median drift of -0.47 %/decade. Table 4.17 summarizes the compliance with user requirements, including also spatiotemporal coverage and resolution.

Table 4.17 - Compliance of the Level-4 MSR total ozone data with user requirements (URD v3.1).

Topic	Requirement	Compliance / evaluation
Horizontal resolution	< 20-100 km	0.5 x 0.5 degree ²
Observation frequency	Daily – weekly	Monthly
Time period	(1980-2010) – (2003-2010)	01/1979 – 12/2019
Total uncertainty	2% (radiative forcing studies)	Bias: 0.2%, spread approx. 2% (includes some co-location mismatch)
	3% (variability studies)	
Stability	1 – 3 % / decade	-0.47% / decade



5 Validation of Nadir Ozone Profile Data Products

5.1 Nadir ozone profile CRDP

The ESA Ozone_cci+ Climate Research Data Package (CRDP) contains twelve nadir ozone profile (NP) products. Table 5.1 lists these products, together with their time range and current availability. All Level-2 (L2) UV-VIS instrument retrievals are performed by the Rutherford Appleton Laboratory (RAL) algorithm, while the thermal infrared measurements of the IASI instruments are processed by a collaboration between the Belgian ULB (Université Libre de Bruxelles) and the French LATMOS (Laboratoire Atmosphères, Milieux, Observations Spatiales, Paris), using their FORLI (Fast Optimal Retrievals on Layers for IASI) algorithm. A joint UV-VIS-NIR retrieval for the GOME-2 and IASI instruments is under development by RAL, while DLR (Deutsche Luft- und Raumfahrt) develops a Level-3 (L3) spatiotemporally gridded product from all RAL's UV-VIS instrument retrievals combined (GOP-ECV).

This version of the Ozone_cci+ PVIR focuses on the validation of the updated (v3) and new (NP_GOME2C) RAL retrieval products and on the updated (version 20191122) and new (NP_IASIC) FORLI products. The validation targets for this PVIR are therefore [RD2]:

- Validation of the L2 nadir ozone profile's updated and new retrieval products.
- Comparison of RAL retrievals from GOME-2C with those of GOME-2A and GOME-2B to confirm the consistency between the three GOME-2/Metop instruments.
- Comparison of the RAL retrieval scheme update (v3) with the previous L2 version (v2) for accuracy improvements.
- Comparison of FORLI-O3 retrievals from IASI-C with those of IASI-A and IASI-B to confirm consistency between the three IASI/Metop instruments.
- Comparison of IASI FORLI retrieval scheme updates (v20191122) for accuracy improvements in the UTLS with the standard FORLI v20151001 processing.

Next to the post-retrieval screening by the data provider, additional filtering criteria have been applied (see Table 5.3). From all approved L2 nadir ozone profile data, only those that are located within 300 km of an NDACC, SHADOZ, or WOUDC ozonesonde or stratospheric lidar station location are retained for further analysis (see Section 5.2). This 300 km radius however is narrowed down for each instrument individually, depending on the instrument's pixel size (see Table 5.2).



Table 5.1 - Overview of Ozone_cci+ nadir ozone profile data products and their current availability: Years marked in blue indicate availability from Ozone_cci Phase II, while purple indicates that the retrieval for that year has been updated or extended in the current Ozone_cci+ project. Crosses mark expected data.

L2 Data Product	Processing entity	Time period																									
		96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
NP_GOME	RAL																										
NP_SCIAMACHY	RAL																										
NP_GOME2A	RAL																										
NP_GOME2B	RAL																										
NP_GOME2C	RAL																										
NP_OMI	RAL																										
NP_TROPOMI	RAL																										
NP_GOME2-IASI	RAL																										
NP_IASIA	ULB/LATMOS																										
NP_IASIB	ULB/LATMOS																										
NP_IASIC	ULB/LATMOS																										
NP_GOP-ECV	DLR	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 5.2 - Overview of Ozone_cci+ L2 nadir ozone profile data product specifications, including local solar time (LST) of the satellite overpass, pixel size, and co-location distance selection based on pixel size.

L2 data product	LST	pixel size (km ²)	Co-location
NP_GOME	10:30AM	320 x 40 km ²	100 km
NP_SCIAMACHY	10:00AM	240 x 32 km ²	100 km
NP_GOME2A/B/C	09:30AM	160 x 160 km ²	100 km
NP_OMI	01:30PM	52 x 48 km ²	50 km
NP_TROPOMI	01:30PM	pixel-adding TBD	TBD
NP_GOME2-IASI	09:30AM	pixel-merging TBD	TBD
NP_IASIA/B/C	09:30AM(+PM)	12 km (diam.)	10 km



Table 5.3 - L2 nadir ozone profile filtering criteria considered in this work (first column) and their settings for the RAL UV-VIS retrieval algorithm (second column) and the FORLI TIR retrieval algorithm (third column). Values that do not comply with the settings are rejected as suggested by the respective data providers.

Filtering criterion	UV-VIS RAL algorithms	TIR FORLI algorithms
Averaging kernel matrix	/	- DFS > 1 - All elements < 2 - First derivative < 0.5 - Second derivative < 1
Chi-square test	1	1
Convergence	1	1
Cost function (normalised)	< 120 (< 2)	/
Effective cloud fraction	< 0.20	< 0.13
Negative ozone values	Rejected	Rejected
Product-specific	- GOME-2A/B/C: January-to-May band 1 SCD < 500 DU - OMI: outer two pixels from each swath rejected - Entire profile rejected upon Band-1A retrieval step failure	- Ozone rejected if incomplete H2O retrieval - IASI-A: rejected from April-September 2015
Solar zenith angle	< 80°	< 83° (day-time) or > 91° (night-time)
Surface pressure	Rejected if unrealistic	Rejected if unrealistic
Surface temperature	/	Rejected if unrealistic
Tropospheric ozone	/	Ratio of 6 km integrated column to total integrated column > 0.085

5.2 Validation approach

The ten-step nadir ozone profile QA/validation chain as applied in this work has already been extensively described within CCI context in [RD46] and [RD47]. Next to data and information content studies, ground-based data records are used as a transfer standard against which the nadir ozone profile retrievals are compared.

5.2.1 Information content studies

Each quantity that is retrieved using the optimal estimation technique contains information both from the satellite measurement and from the a-priori profile and covariance matrix. The contribution of prior information can be significant where the measurement is weakly or even not sensitive to the atmospheric ozone profile, e.g. in case of fine-scale structures of the profile, below optically thick tropospheric clouds, and at the lower altitudes. The information distribution is captured by the retrieval's ex-ante vertical averaging kernel matrix A (sometimes also AKM hereafter), which represents the sensitivity of the retrieved state \hat{x} to changes in the true profile x_t at a given altitude: $A(m, n) = \partial \hat{x}(m) / \partial x_t(n)$.

A study of the algebraic properties of this averaging kernel matrix, denoted information content study, can help understanding how the system captures actual atmospheric signals. Through straightforward analysis however, it can be easily demonstrated that typical information content measures as discussed in this section usually depend on the units of the averaging kernel matrices they are calculated from [RD46]. As these measures however should be unit-independent, fractional AKMs A_F must be considered.



Starting from the averaging kernels provided as part of the Ozone_cci+ CRDP L2 nadir ozone profile products, the degree of freedom in the signal (DFS) and the vertical sensitivity are studied. These quantities are given by the AKM trace and row sum profile, respectively. The DFS of a retrieved atmospheric profile is a non-linear measure for the number of independent quantities that can be determined and as such loosely related to the Shannon information content [RD73]. The vertical sensitivity to the measurement is a unit-normalised measure for how sensitive the retrieved ozone value at a certain height is to ozone values at all heights.

Besides the more common DFS and sensitivity information content quantities, in this work the vertical averaging kernels' offset and width are considered as well. The offset is an estimate of the uncertainty on the retrieval height registration, given here by the direct vertical distance (in km) between an averaging kernel's peak sensitivity altitude z_{peak} and its nominal retrieval altitude z_{nom} as $d(m) = z_{peak}(m) - z_{nom}(m)$. Ideally, within each kernel, this distance equals zero. Ozone_cci+ user requirements also specify an upper limit for the vertical resolution of the nadir ozone profile retrievals. Several methods have been proposed to estimate the vertical resolution from the width of the vertical averaging kernels (see overview in [RD46]), but usually it is determined either as a full width at half-maximum (FWHM) value around the kernel's peak altitude or as the Backus-Gilbert spread (BG) or resolving length around its centroid.

5.2.2 FRM comparisons

The ground-based FRM data considered for the nadir ozone profile validation in this report has been collected from ESA's Atmospheric Validation Data Centre (EVDC). The EVDC Cal/Val data portal contains ozonesonde data from the NDACC, SHADOZ, and WOUDC network archives, and additionally collects ozonesonde data in near-real time within the MATCH campaign (<https://evdc.esa.int/campaigns/o3sondes/>). Stratospheric lidar data originate from the NDACC data archive and its rapid delivery section. The EVDC data portal redistributes the ground-based data in a harmonized HDF5 GEOMS format.

Like for the satellite data, prior to searching for co-locations with satellite ECV data, data screening has been applied to ground-based correlative measurements by ozonesondes and lidars, both on entire profiles and on individual altitude levels. The recommendations of the ground-based data providers to discard unreliable measurements are followed. Measurements with unrealistic pressure, temperature, or ozone readings are rejected automatically. Ozonesonde measurements at pressures below 5 hPa (beyond 30-33 km) and lidar measurements outside of the 15-47 km vertical range are rejected automatically as well. Prior to these data manipulations, the ground-based ozone profile data were converted to partial ozone column units (DU) by vertical integration. While ozonesondes report measurements in partial pressure, easily converted into VMR units (ppmv) and in ND using the on-board PTU measurements, the lidar data are given in number density.

Only co-locations with a maximal spatial distance of 100 km or smaller (see Table 5.2) and a maximal time difference of one day were allowed. When multiple satellite pixel co-locations with one unique ground measurement occur, only the closest satellite measurement is kept. Calculating difference profiles requires harmonisation of the satellite retrieval and ground-based reference ozone profiles in terms of at least their representation and vertical sampling. In order to down-sample a ground-based ozone profile measurement to the satellite retrieval grid, a mass-conserving regridding in subcolumn units on layers is preferably used [RD60]. This technique however has been extended to be applicable in number density or volume-mixing ratio units on levels as well (also see next sections). Additionally, the satellite and ground-based profiles' vertical smoothing difference error is minimized by averaging kernel multiplication [RD48, RD73].

The baseline output of the L2 validation exercises consists of median absolute and relative nadir ozone profile differences at individual stations or within latitude bands for the entire time series. This median difference is a robust (against outliers) estimator of the vertically dependent systematic error, i.e. the bias, of the satellite



data product. The bias profiles for the entire list of stations are then combined and visualized as a function of several influence quantities in order to reveal any dependences of the systematic error. The influence quantities considered in this work are latitude (meridian dependence), total and tropospheric ozone column, DFS, SZA (solar zenith angle), VZA (viewing zenith angle), (effective) cloud fraction (for the UV-VIS products), and thermal contrast (for the TIR products).

Besides the median difference, also the Q84-Q16 interpercentile (IP68) of the differences is calculated as a robust spread estimator of the random errors in the satellite data product, i.e. the precision profile. However, this spread on the differences will also include contributions from ground-based random uncertainties (limited to a few percent) and representativeness (sampling and smoothing) differences between the satellite and reference measurements, and therefore in fact provides an upper limit on the actual random satellite uncertainty. In case of a normal distribution of the ozone differences, median and IP68 are equivalent to mean and standard deviation, but they offer the advantage to be much less sensitive to occasional outliers.

5.3 Validation results

5.3.1 Consistency of GOME-2A/B/C instrument retrievals

RAL L2 v3 nadir ozone profile retrievals are now available for all three GOME-2 instruments. This allows performing a GOME-2A/B/C retrieval consistency check, which is ideally performed using identical processor versions, prior information, time ranges, and ground-based reference data for all three instruments. This has however not been achieved in practice. Processor versions and retrieval periods slightly differ, as indicated in Table 5.4 below. As a result, the ground-based FRM data differs between the three instruments as well. The presented comparison results therefore provide an indicative assessment only, and require confirmation upon (re)processing using exactly the same processor settings and for identical time intervals.

Table 5.4 - Overview of the latest L1 and L2 processor versions and prior source per instrument for the nadir ozone profiles delivered by RAL. Next to the start and end months of the full datasets, the years considered for the Metop-A/B/C consistency check (this section) and the delta-validation (next section) are indicated.

Instrument	L1	RAL L2	Prior source	Start	End	Consistency check	Delta-val. w.r.t. L2 v2
GOME	GDP v4	v0301	ERA-I	1995/06	2011/06	/	1996-2010
SCIAMACHY	v7.04	v0300	ERA-I	2002/08	2012/04	/	2003-2010
GOME-2A	v6.0-6.3	v0300	ERA-I	2007/01	2019/08	/	2008-2019
GOME-2A	v6.3-7.0	v0303	ERA-5	2019/09	2021/12	2020-2021	/
GOME-2B	v6.3	v0303	ERA-5	2014/06	2020/11	/	2014-2020
GOME-2B	v6.3-7.0	v0305	ERA-5	2020/11	2021/11	2020-2021	/
GOME-2C	v6.3	v0300	ERA-5	2020/01	2021/06	2020-2021	/
OMI	v003	v0214	ERA-I	2004/10	2019/08	/	/
OMI	v003	v0214e5	ERA-5	2019/09	2021/10	/	/

Looking at the comparisons in Figure 5.1 (see also Figure 5.2 and Figure 5.3), which involve averaging kernel smoothed satellite profiles, one observes that generally the RAL v3 GOME-2A/B/C retrieval products agree similarly with the ground-based data, showing a rather typical Z-curve with zero biases approximately at 5-10 and around 20 km altitude. GOME-2B and C show a negative bias peak in the UTLS (5 to 20 km) and a positive bias peak in the upper stratosphere (between 20 and 55 km) that both amount to about 20 to 40 % (with a higher negative peak for GOME-2B). The stratospheric lidar comparisons on the other hand show a negative stratospheric bias of 10-20 % for GOME-2A. For all three instruments, the bias again shifts towards



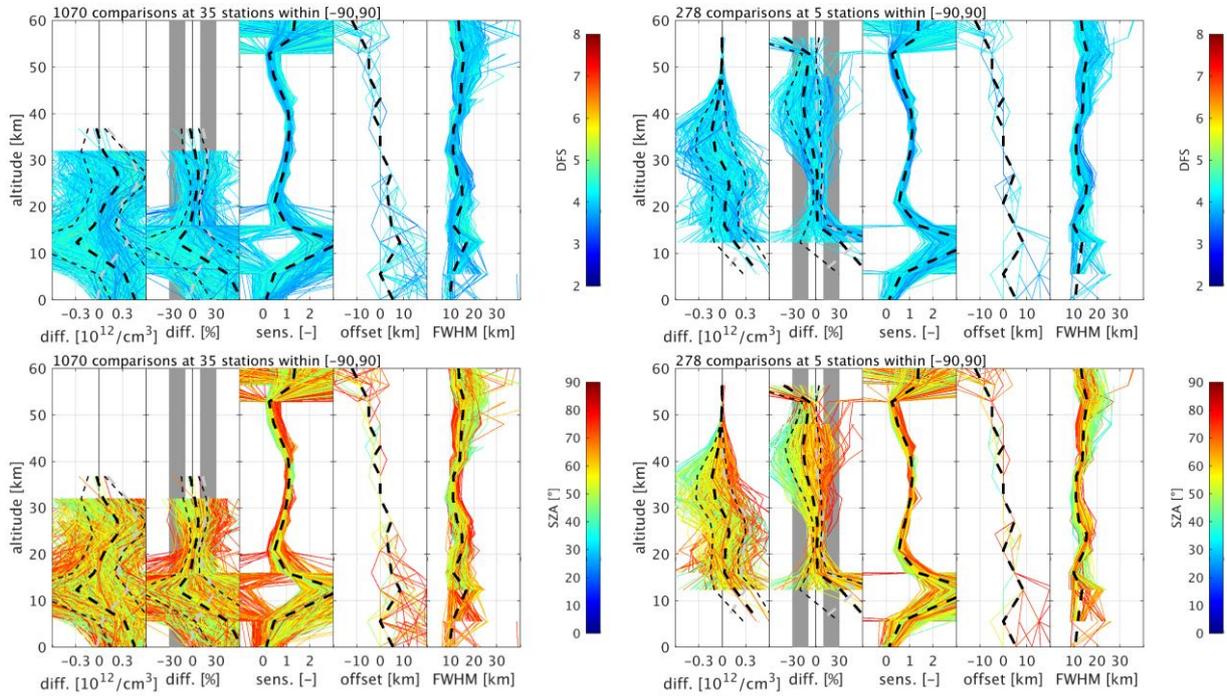
positive values below 5 km, with a high positive bias towards the surface for GOME-2A. The sensitivity for this lowest layer however is reduced to about 0.5 or below, meaning that 50 % or more of the retrieval information comes from the prior profile rather than from the measurement.

The somewhat deviating bias behaviour for GOME-2A with its strong positive tropospheric bias could be at least partially attributed to instrument degradation, especially for the recent years under consideration. This is also clear from the mean number of DFS that is obtained for the GOME-2 instruments, which roughly amount to 4, 4.5, and 5 for GOME-2A, B, and C, respectively. Apart from some seasonality (up to 0.5 at maximum), these DFS values change little for each instrument. This is quite remarkable, given the occurrence of negative sensitivities in the UTLS for all three instruments, which mostly correspond to high solar zenith angle observations (above about 65 degrees, see Figure 5.1). From the ozonesonde comparisons, it becomes clear that the observations showing negative UTLS sensitivities correspond with the highest negative UTLS biases. On the other hand, these observations also induce the highest stratospheric biases, as can be seen from the lidar comparisons, which show a very clear SZA dependence for all three GOME-2 instruments. It could therefore be appropriate to strengthen the SZA screening from say 80° to 70° and/or introduce a screening of negative sensitivities. This would reduce both the positive (stratosphere) and negative (UTLS) biases along the vertical profile.

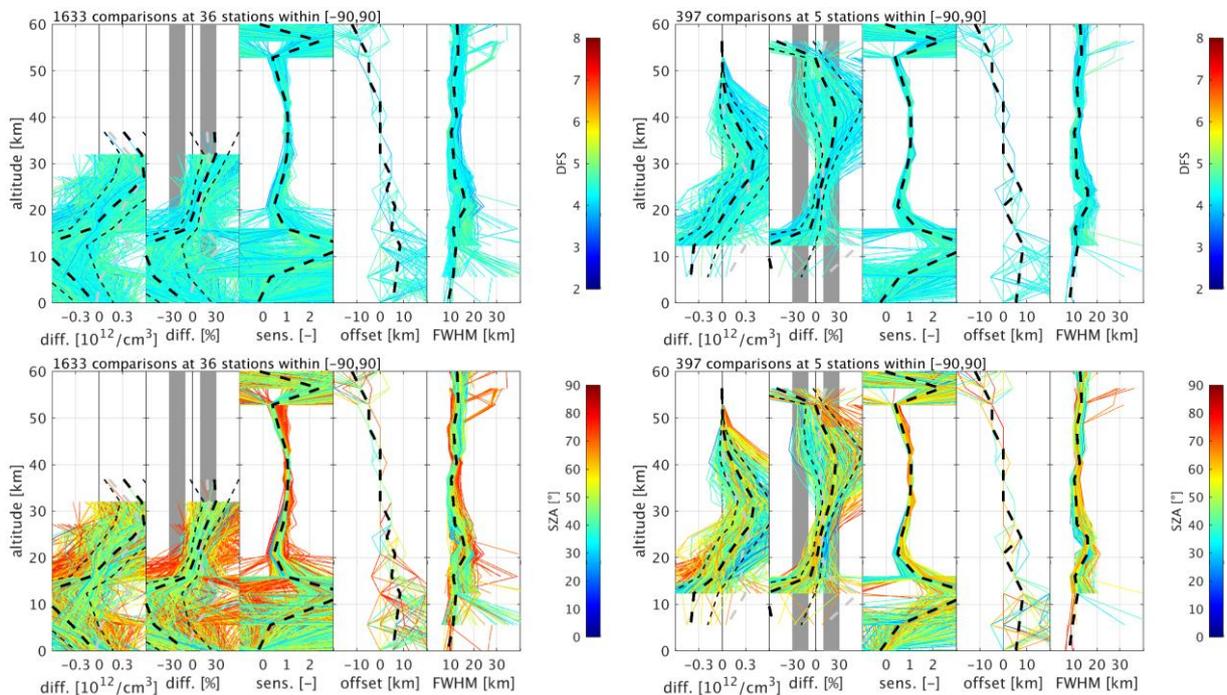
For all three GOME-2 instruments, the comparison uncertainties in terms of the 68 % interpercentile spread display a U-shaped curve with a minimum of about 10 % around 20-25 km (plotted as dashed lines around the median difference in Figure 5.1). This dispersion increases to roughly 30 % at 45 km, to slightly decrease again above, but rises even more strongly in the UTLS where the sensitivity profile peaks and towards the ground. The bias discussed above exceeds this dispersion (becomes significant), in the troposphere for GOME-2A, and in the UTLS and stratosphere for GOME-2B and C.



GOME-2A v3.03 2020/01-2021/12



GOME-2B v3.05 2020/11-2021/11



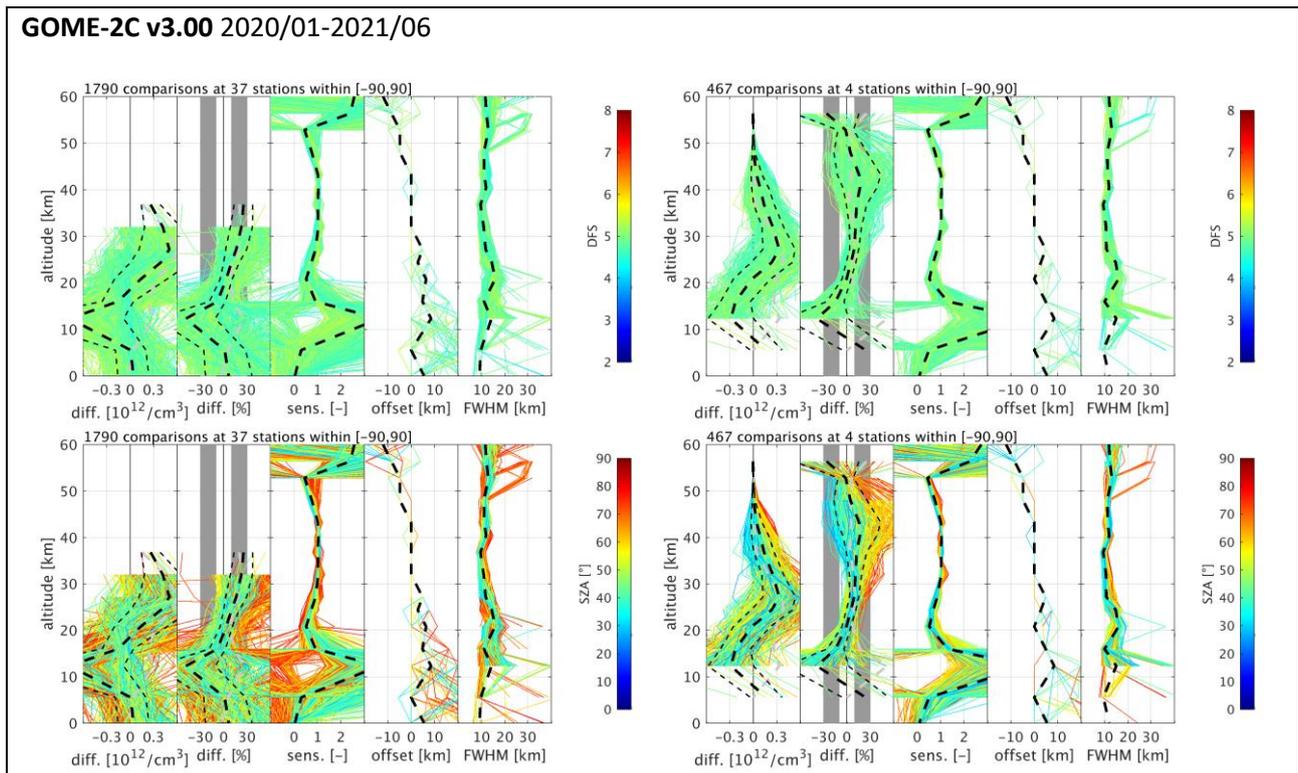
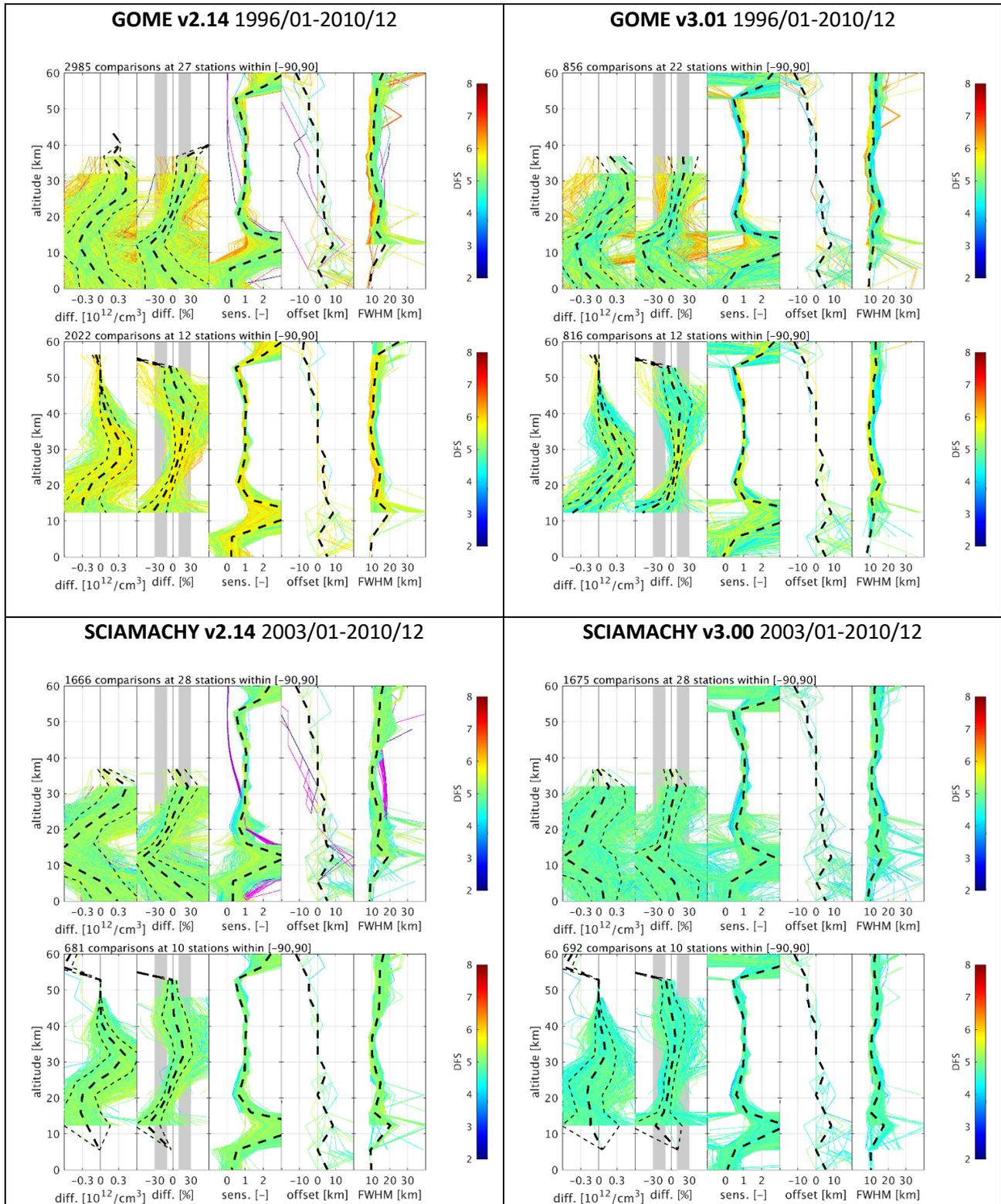


Figure 5.1 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of RAL v3 retrievals for Metop GOME-2A/B/C (top to bottom, respectively) with ozonesonde (left) and lidar (right) reference measurements (the exact processor version and time range is indicated with each instrument). Individual profile statistics are plotted as a function of DFS and solar zenith angle (SZA), using the colour coding indicated by the colour bar on the right of each plot.

5.3.2 Assessment of the RAL v2 to v3 retrieval updates

The RAL v2 UV-VIS nadir ozone profile retrievals were extensively discussed in [RD47]. The corresponding comparison results are again plotted as a function of DFS in Figure 5.2, left column. The right column contains the ozonesonde and lidar comparison results for v3 of the RAL retrieval. Note that time ranges have been extended for the GOME-2A and B instruments, while the number of comparisons is much reduced for GOME because of more severe screening settings. Nevertheless, it is quite clear that with the update from RAL v2 to v3, the satellite nadir ozone profile uncertainties have hardly changed. The exception is SCIAMACHY, which shows slightly reduced uncertainties in terms of both bias and dispersion for v3, both in the troposphere and stratosphere.

On the other hand, it is remarkable that the average retrieval DFS has decreased by up to 0.5 in general. This is most pronounced for GOME-2A and B, although one has to take into account that their time series extension including increased instrument degradation plays a role as well. The overall DFS degradation seems to go hand in hand with the appearance of strongly negative UTLS sensitivities in v3 of the RAL retrieval, which were not present in its v2 (negative sensitivities were present in the lower troposphere in v2 already). This observation confirms that the introduction of a screening of ozone profiles with negative sensitivity in the UTLS could improve the overall performance, i.e., reduce the average uncertainty, of the RAL v3 products, as already indicated in the previous section.



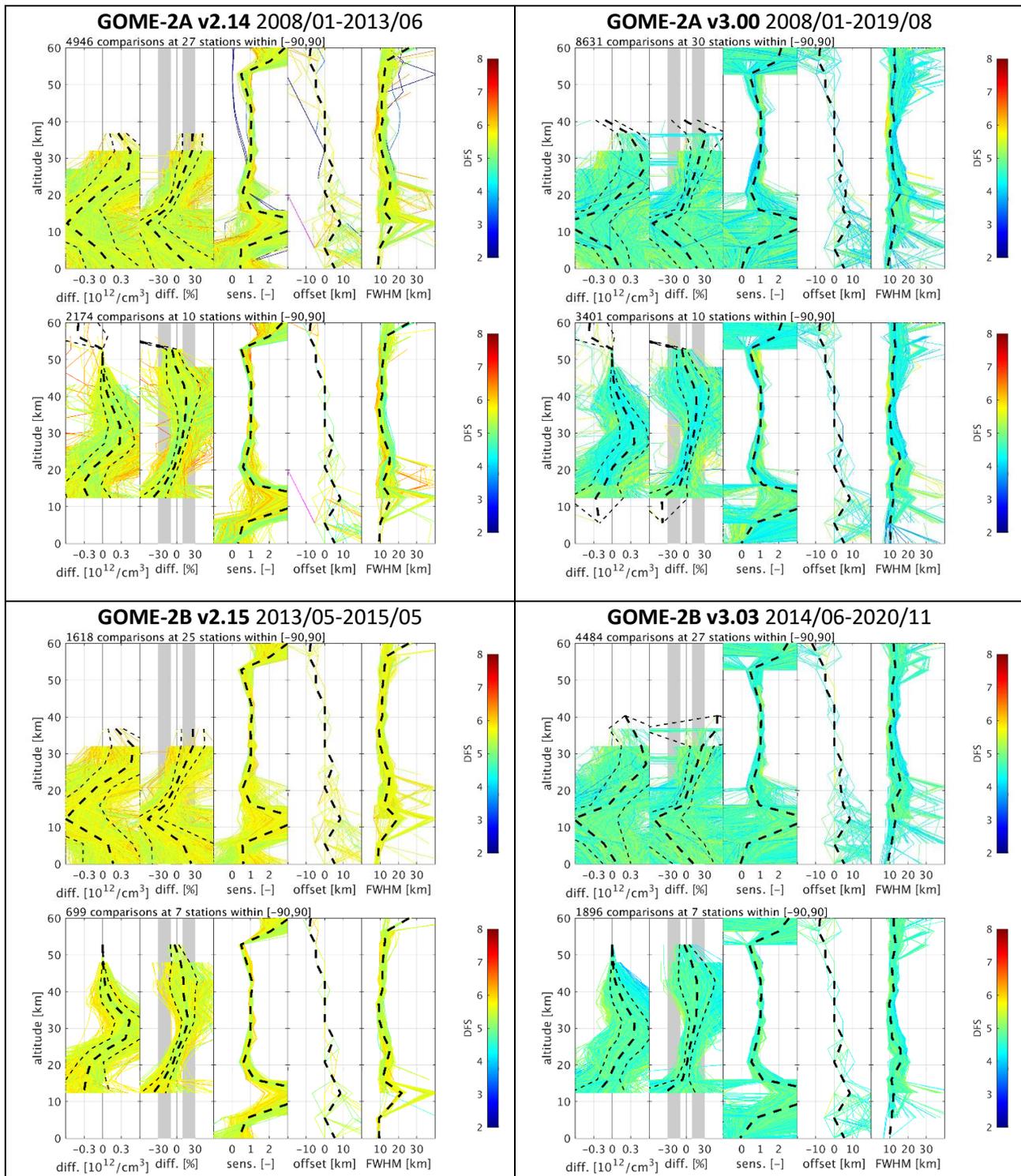


Figure 5.2 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of RAL v2 (left) and v3 (right) retrievals for GOME, SCIAMACHY, and GOME-2A/B (top to bottom, respectively) with ozonesonde and lidar reference measurements (the exact processor version and time range is indicated with each instrument). Individual profile statistics are plotted as a function of DFS using the colour coding indicated by the colour bar on the right of each plot.

5.3.3 Validation of the extended OMI time series

Also for the OMI instrument, the RAL v2 nadir ozone profile retrieval validation was extensively discussed in [RD47], up to 2015. Figure 5.3 displays information content and comparisons with respect to ozonesonde and lidar measurements for the extended OMI time series, up to August 2019, again as a function of DFS and SZA. These results are qualitatively very similar to those previously reported, and to those of the other UV-VIS instrument retrievals by RAL (v2) as discussed above. E.g., the solar zenith angle dependence of the (mostly stratospheric) bias is clearly visible, with the bias again being highest for the highest SZA retrievals. On the other hand, two important differences occur. First, with values of 6 to 6.5 on average, the OMI retrieval typically contains more degrees of freedom than the other UV-VIS nadir ozone profile products (a few profiles with nearly zero sensitivity in the stratosphere and hence a DFS below 4 could be additionally screened). Second, relative differences remain limited in both the UTLS and the stratosphere, with average values being smaller than the dispersion and within the user requirements overall (i.e., below 30 %, as indicated by grey boxes in the plots).

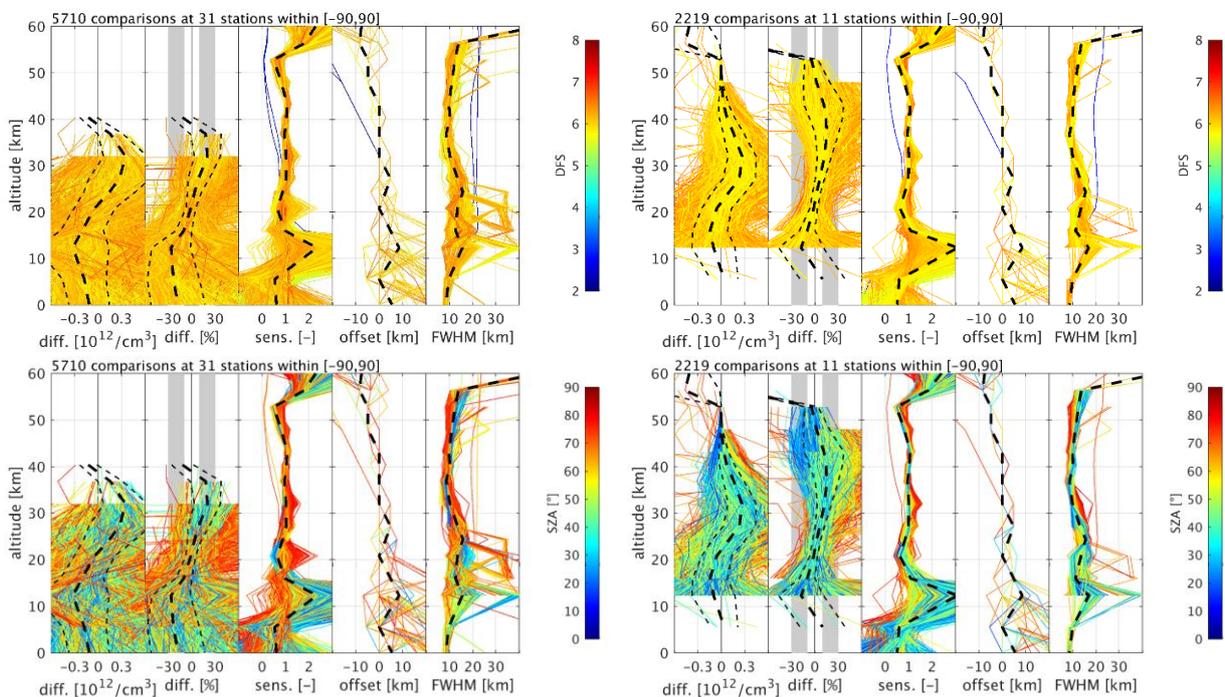


Figure 5.3 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of RAL retrievals for OMI with ozonesonde (left) and lidar (right) reference measurements (processor version 2.14 for 2005/01 to 2019/08). Individual profile statistics are plotted as a function of DFS (top) and solar zenith angle (SZA, bottom) using the colour coding indicated by the colour bar on the right of each plot.

5.3.4 Consistency of IASI-A/B/C instrument retrievals

Figure 5.4 to Figure 5.7 contain the median (relative) differences, 68% interpercentile spreads, vertical sensitivities, offsets, and effective vertical resolutions (FWHMs) for the comparison of FORLI v20151001 and v20191122 retrieved IASI profiles with ground-based reference measurements for DFS and thermal contrast as influence quantities (others are not shown in this report). IASI-A/B/C retrievals using FORLI v20151001 are always shown on the left for data from October and November 2019, while FORLI v20191122 data are shown on the right for January to February 2020 (as the version switch on the data server was made mid-December 2019). Note that for IASI/Metop-A November 13-17 data is missing.



The IASI-A/B/C retrieval consistency check is ideally performed using identical ground-based reference data for all three instruments. In practice however, due to slightly different station overpass characteristics and satellite instrument operations, this was not fully achieved, although most collocations for all three satellite instruments are for the same FRM data. Analogously, the assessment of the FORLI update v20191122 with respect to the previous v20151001 is ideally based on retrievals from exactly the same Level-1 data. This is not feasible from the operational data considered here, which are taken from a server with a hard version switch, but could be achieved in the future for a test dataset of double retrievals.

Figure 5.4 to Figure 5.7 demonstrate that the retrieval results for all three IASI instruments are very similar, showing no significant differences between their respective statistics, and this for both FORLI retrieval versions under consideration. The IASI instruments show a less than 10 % and insignificant stratospheric bias, a 10 to 30 % insignificant positive bias in the UTLS, and an order of 10-20 % negative bias at the edge of significance in the troposphere. The latter is in agreement with a tropospheric ozone (from IASI-A and IASI-B retrieved with FORLI v20151001) validation exercise performed by Boynard et al. [RD29]. Possible reasons for the positive UTLS bias are discussed in [RD35]. The comparison results show hardly any scan angle (VZA) dependence or seasonality, except for some larger systematic differences around the Antarctic ozone hole that can be partially attributed to co-location errors at the edge of the polar vortex (not shown here). The remaining meridian dependences are typically limited to stronger UTLS bias fluctuations in the tropics.

The vertical sensitivity profiles are close to unity around the ozone peak and above (25 to 35 km) for all three instruments. Typically, the sensitivity decreases above and below due to the smaller ozone concentrations. The FORLI retrievals show sensitivity fluctuations around the UTLS, ranging between 0 and 2. Although the overall IASI sensitivity variability is strongest around the equator, these outliers typically occur in the polar regions and go together with excessively high retrieved ozone peaks. The strong sensitivity variability in general hampers the averaging kernel smoothing of the reference profiles before comparison, as this procedure then introduces a bias instead of reducing the vertical smoothing difference error. Usually however, except for decreased surface-level sensitivity (0.5) and a median 1.5 peak around the UTLS with slight compensation above and below, the FORLI sensitivity is more vertically consistent.

The retrieval offset (vertical registration uncertainty) amounts to about 5-10 km on average, but shows some discreteness due to the FORLI retrievals being performed on a fixed 1 km vertical grid. Moreover, the offset shows specific features for all three IASI instruments, like the peaks at 5 and 15 km, and a jump near 25 km altitude. The tropospheric offset sometimes even explodes to unrealistic values, when the corresponding averaging kernels have no clear maximum. The dependence on DFS and thermal contrast however is rather small. The behaviour of an averaging kernel's sensitivity and offset is typically also reflected in its width, which is here measured by the kernel's FWHM. This FWHM ranges between 10 and 15 km on average for the IASI instrument retrievals, with oscillations occurring in the UTLS that are again little dependent on DFS or thermal contrast.

5.3.5 Assessment of FORLI retrieval update to v20191122

Looking at the FORLI v20151001 comparison results into detail (also see [RD71]), it becomes clear that both the polar sub-tropopause and the tropical stratosphere and UTLS difference outliers seem to go together with a thermal contrast dependence of the differences (clearer for the lidar comparisons) that also agrees with the sensitivity dependence. One would expect the thermal contrast to be mainly influential in the lowermost layers, but the information content studies on the IASI product have demonstrated that the corresponding averaging kernels show significant oscillations that are vertically interdependent. Therefore, the sensitivity outliers around 30 km altitude are related to the strongly negative thermal contrasts (order of minus 10 K or below) mostly occurring for nighttime measurements (not shown here) and typically go together with very low DFS values and strong ozone over-estimations. Minimal retrieval sensitivities and DFS



values appear to occur when these modes are combined with a high scan angle (VZA) observation (also not shown). On the other hand, highly positive thermal contrasts, order of 10 K or above, sometimes occurring in combination with high solar zenith angles, yield increased positive UTLS biases.

The FORLI v20191122 retrievals in fact show similar features and dependences for all three IASI instruments, but two important observations can be made in comparing both retrieval processor versions. First, it is quite clear that the seasonal and meridian DFS and information content variation is reduced in the last retrieval version, i.e. typical DFS values are closer to 3 instead of showing strong fluctuations. A downside of this might be that the overall median DFS has slightly decreased (to be confirmed with longer data time series), but it increases the global retrieval consistency. Correspondingly, the kernel FWHM fluctuations around the 10 km average in the UTLS for FORLI v20151001 have increased to about 15 km on average in v20191122, which, given the sensitivity fluctuations and positive biases within this vertical range, looks more realistic from the information content perspective.

Second, and relatedly, the more consistent FORLI v20191122 retrieval appears to result in a reduced comparison spread as an estimator of the retrieval's random uncertainty. This is most clear in the UTLS for the ozonesonde comparisons, and in the stratosphere for the lidar comparisons. The tropospheric random uncertainty looks unaltered. In addition, the vertical extent and absolute value of the positive bias around the UTLS appears to have decreased for the latest processor version. Both effects yield a slightly reduced combined uncertainty for the FORLI v20191122 algorithm, but this is again to be confirmed when more data are analysed.



IASI-A/B/C v20151001 vs. ozonesonde

IASI-A/B/C v20191122 vs. ozonesonde

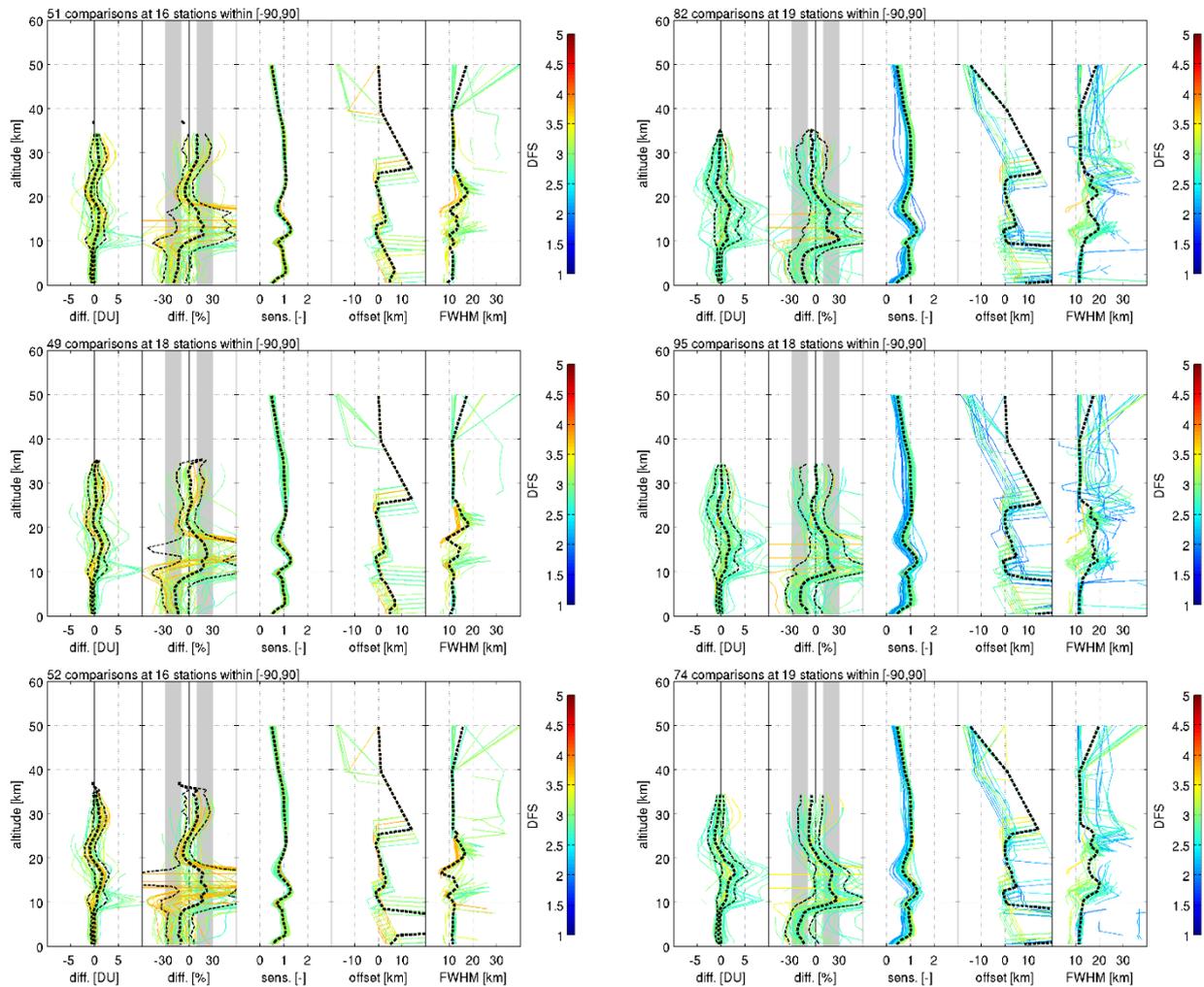


Figure 5.4 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of FORLI v20151001 (left) and v20191122 (right) L2 IASI-A/B/C (top to bottom, respectively) retrieved profiles with ozonesonde reference measurements (2019-2020). Individual profile statistics are plotted as function of DFS using the colour coding indicated by the colour bar on the right of each plot.

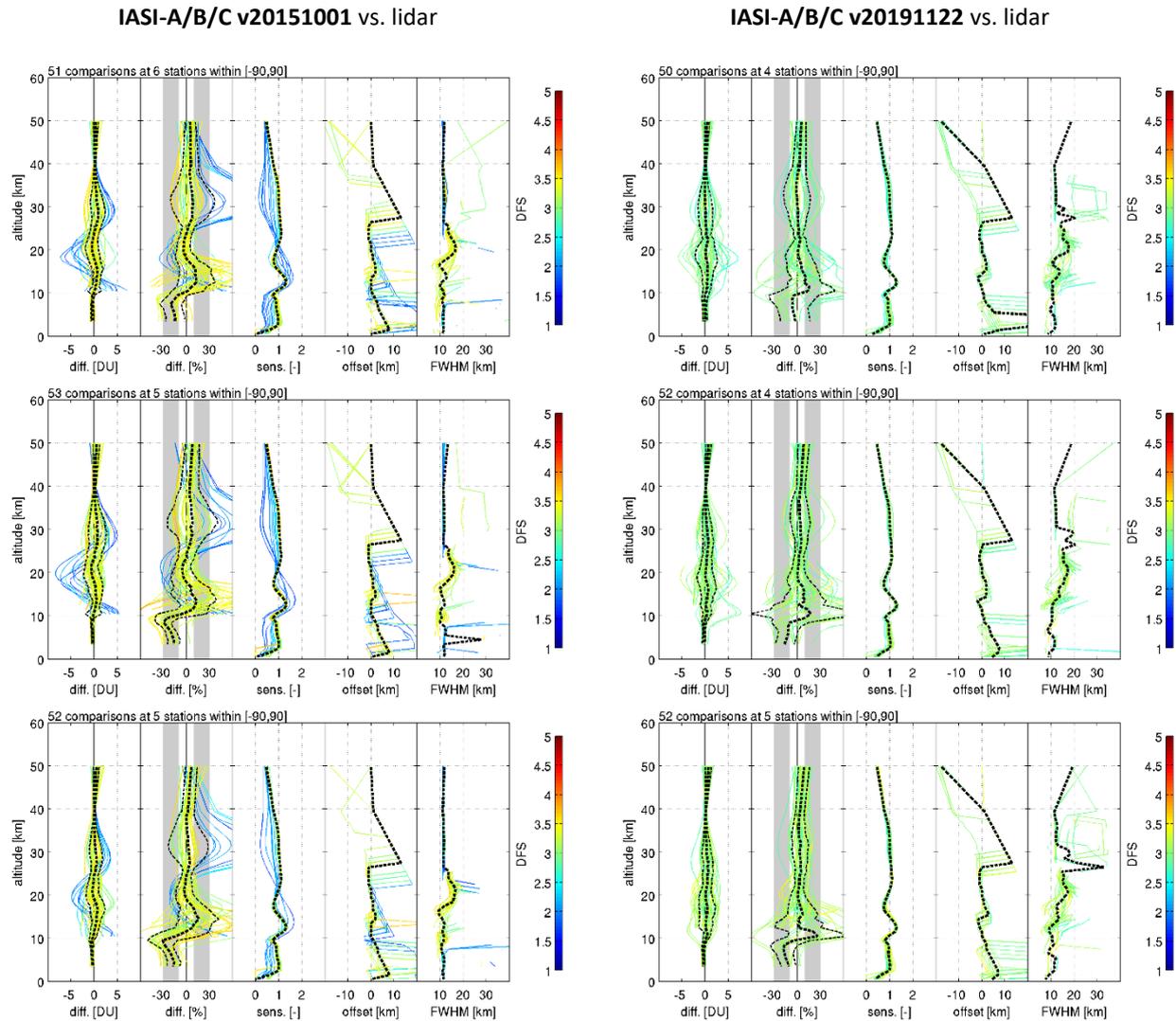


Figure 5.5 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of FORLI v20151001 (left) and v20191122 (right) L2 IASI-A/B/C (top to bottom, respectively) retrieved profiles with lidar reference measurements (2019-2020). Individual profile statistics are plotted as function of DFS using the colour coding indicated by the colour bar on the right of each plot.



IASI-A/B/C v20151001 vs. ozonesonde

IASI-A/B/C v20191122 vs. ozonesonde

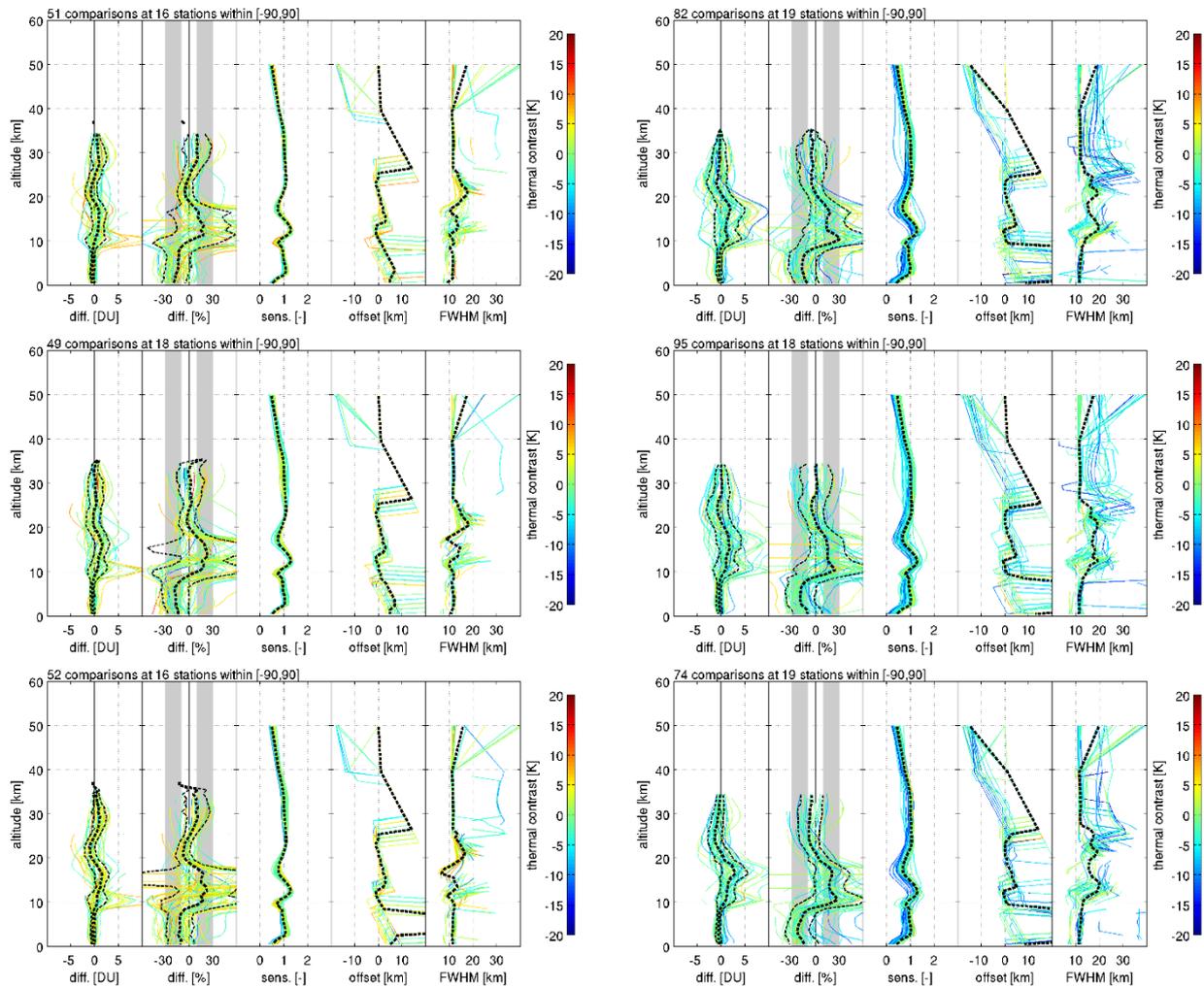
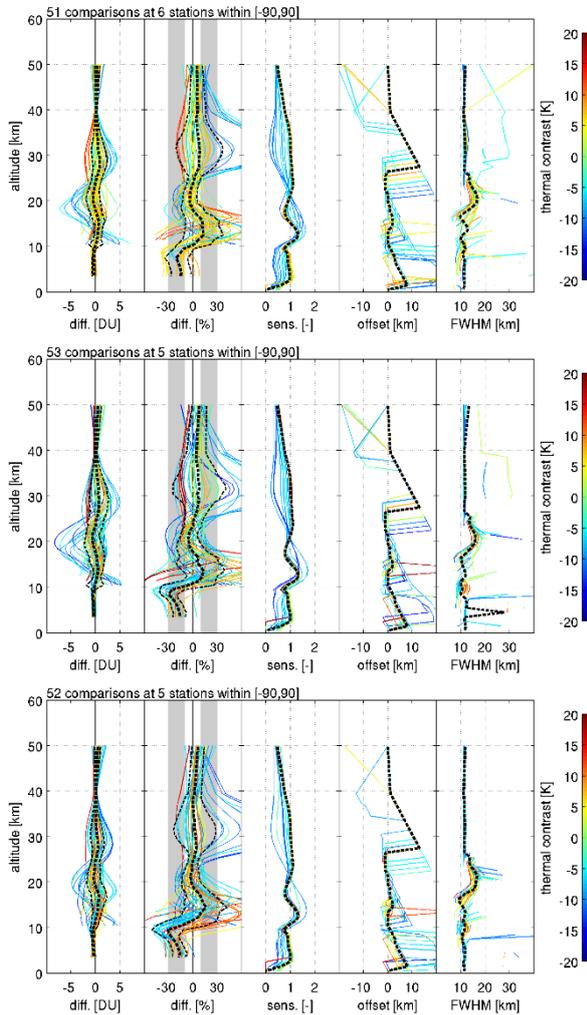


Figure 5.6 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of FORLI v20151001 (left) and v20191122 (right) L2 IASI-A/B/C (top to bottom, respectively) retrieved profiles with ozonesonde reference measurements (2019-2020). Individual profile statistics are plotted as function of thermal contrast [K] using the colour coding indicated by the colour bar on the right of each plot.



IASI-A/B/C v20151001 vs. lidar



IASI-A/B/C v20191122 vs. lidar

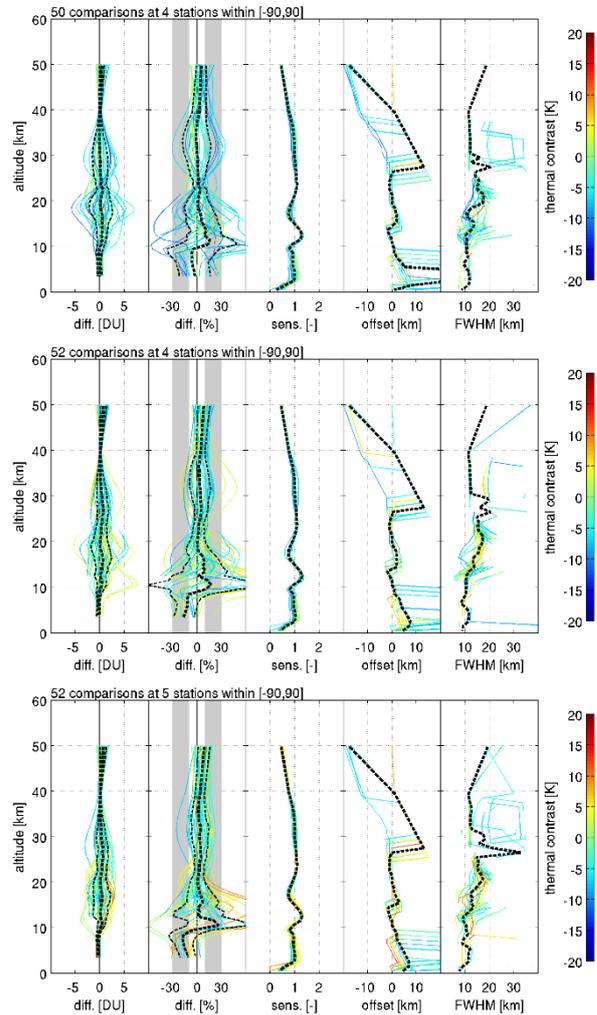


Figure 5.7 - Median absolute and relative differences (thick dashed lines), 68% interpercentile spreads (thin dashed lines), and median vertical sensitivities, offsets and FWHMs (dashed lines) for comparison of FORLI v20151001 (left) and v20191122 (right) L2 IASI-A/B/C (top to bottom, respectively) retrieved profiles with lidar reference measurements (2019-2020). Individual profile statistics are plotted as function of thermal contrast [K] using the colour coding indicated by the colour bar on the right of each plot.



5.4 Results discussion and conclusions

Table 5.5 collects the major QA/validation quantities discussed throughout this chapter, their corresponding typical values as discussed in the previous sections, and indications of their compliance with the GCOS user requirements. The nadir ozone profile products under study cover the 1995 to 2021 time window globally, which is sufficiently long for (drift-corrected) ozone trend studies according to the GCOS user requirements (UR). They also fulfil the GCOS user requirements in terms of observation frequency and horizontal and vertical resolution. Only for the latter one has to keep in mind that the L2 nadir products show UTLS sensitivity outliers and are strongly correlated vertically due to averaging kernel fluctuations that extend beyond the kernel's 10-15 km FWHM.

The Ozone_cci+ nadir ozone profile products under study typically do not comply with the GCOS user requirements in terms of total uncertainty. The total uncertainty is thereby determined as the quadratic sum of the products' systematic and random uncertainties, which on their turn are estimated from the comparison (with ground-based reference measurement) bias and dispersion, respectively. Whereas the RAL v2 and v3 UV-VIS retrieved products show a typical Z-curve bias with strong 20-40 % positive (stratosphere) and negative (UTLS) maxima, the FORLI retrievals' systematic uncertainty is rather consistently of the order of 10 % in the stratosphere and troposphere, but shows stronger fluctuations (20 to 30 %) in the UTLS. Total uncertainties therefore range from about 10 % at minimum in the stratosphere to at least 20 % in the troposphere, and even higher values in the UTLS for IASI and for the UV-VIS instruments.

The nadir ozone profile retrieval performance in terms of sensitivity and systematic and random uncertainties typically depends on the SZA for the UV-VIS retrievals and on the thermal contrast for the IR retrievals. Not shown in this report, as more or less in line with nadir ozone profile retrieval expectations, are that the comparison results depend on the surface albedo and effective cloud fraction (ECF), mainly in the troposphere. The ECF and surface albedo dependence, however, is also reflected, yet inversely, in the UTLS, due to the sensitivity peak in this region. Finally, the quarter and scan pixel index have hardly any effect on the UV-VIS comparison results, meaning that the RAL retrieval algorithm copes with ozone seasonality and instrument viewing angle effects appropriately.

Applying bias corrections to the nadir ozone profile CRDP presented in this work might not yield optimal results however. Next to the L2 data screening recommended by the data providers (summarised in Table 5.3) the validation results presented in the previous sections point at additional data screening options. The bias outliers for the IASI retrievals in the polar troposphere and the tropical UTLS go together with a thermal contrast and sensitivity dependence of the differences. These profiles could therefore be excluded from any further use by insertion of a strongly negative thermal contrast or low DFS value screening, e.g. shifting the DFS screening threshold from one (as suggested by the ULB/LATMOS team) to two. Vertically resolved profile screening could additionally reject consistent altitude-dependent bias outliers. For the UV-VIS nadir ozone profile products retrieved by RAL, it could be appropriate to strengthen the SZA screening from say 80° to 70° and/or introduce a screening of negative sensitivities. This would reduce both the positive (stratosphere) and negative (UTLS) bias on average.



Table 5.5 - Major QA/validation quantities, their corresponding typical values, and indication of GCOS user requirement (UR) compliance for the Ozone_cci+ nadir ozone profile products.

QA quantity (GCOS UR)	IASI-A/B/C TIR retrieval	RAL UV-VIS retrieval
Time period (1996-2010)	2008-2021 (2019-2020 for this report)	1995-2021
L2 observation frequency (daily to weekly)	Both day-time and night-time daily	Global coverage within 3 days
Horizontal resolution (20-200 km)	12 km	32 to 160 km along track, 52 to 320 km across
Vertical resolution (6 km to troposphere)	Fixed 1 km grid but 10-15 km kernel width and strong UTLS fluctuations	Fixed grid with up to 6 km layers but ~15 km kernel width and SZA dep. tropospheric fluctuations
DFS	2-4 with meridian and seasonal dep.	4 to 5.5 with 0.5 seasonality
Vertical sensitivity	Outliers around UTLS	UTLS peak ~3 with under-sensitivity right above and below
Height registration uncertainty / retrieval offset	5-10 km on average, but strong features	< 10 km
Systematic uncertainty estimated from comp. bias	< 10 % stratospheric bias, 20-30 % pos. (UTLS) to ~10-20 % neg. (troposphere)	Z-curve with maxima at 20-40 % pos. (stratosphere) and neg. (UTLS)
Random uncertainty estimated from comp. spread	~10-30 %, slightly reduced for v20191122	10 % around 20-25 km, with higher values above and below
Total uncertainty (16 % below 20 km, 8 % above 20 km)	~10 % stratosphere, 20 % in troposphere, higher in UTLS	10 % minimum at 20-25 km, increasing above and below
Dependence on influence quantities	Thermal contrast especially in polar troposphere and tropical UTLS, agrees with DFS/sensitivity dependence	Higher SZA corresponds to higher biases; small surface albedo and ECF dep. propagates to higher altitudes
Stability (1-3 %/dec.)	(not addressed here)	(not addressed here)



6 Validation of Limb Ozone Profile Data Products

6.1 Introduction

This section reports on the assessment of the Level-2 and Level-3 limb ozone profile datasets of the Ozone_cci+ Climate Research Data Package (CRDP), freely accessible from <https://climate.esa.int/en/projects/ozone/data>.

The Level-2 CRDP consists of a HARMONized dataset of OZone profiles (also named HARMOZ) from the following list of limb/occultation sensors: GOMOS, MIPAS and SCIAMACHY on Envisat, OSIRIS on Odin, ACE-FTS on SciSat-1, OMPS-LP on Suomi-NPP, SAGE II on ERBS, HALOE on UARS, SABER on TIMED, MLS on EOS-Aura, POAM III on SPOT-4, SAGE III on Meteor-3M and SAGE III on ISS. Each HARMOZ data set is screened for outliers by the instrument experts, presented on an identical vertical grid (altitude, pressure, or both) and archived in the same NetCDF format. A description of Level-2 data format, instruments and retrieval algorithms can be found in the README file, the peer-reviewed description of HARMOZ in ESSD [RD80] and the Algorithm Theoretical Baseline Document [RD4].

The Level-3 CRDP is derived from the Level-2 CRDP and consists of (a) monthly mean ozone anomalies in 10° latitude zones for each individual instrument in its native vertical coordinate, (b) monthly mean ozone anomalies in 10° latitude zones combining the data from SAGE II, OSIRIS, GOMOS, MIPAS, SCIAMACHY, ACE-FTS and OMPS-LP, (c) monthly mean ozone anomalies in 10° latitude by 20° longitude cells combining the data from OSIRIS, GOMOS, MIPAS and SCIAMACHY. A detailed description of Level-3 data format and the merging algorithm can be found in the Algorithm Theoretical Baseline Document [RD4] and the README files.

6.2 Harmonized validation methodology

The validation of the HARMOZ ozone profile datasets is primarily based on comparisons with respect to correlative ground-based measurements. The Level-2 validation methodology is described extensively by Hubert et al. [RD43], any deviations will be motivated below. The Level-3 methodology is documented in Section 6.5.1. Ground-based observations are acquired at a variety of ozonesonde, lidar and microwave radiometer (MWR) stations performing network operation within WMO's Global Atmosphere Watch (GAW), the Network for the Detection of Atmospheric Composition Change (NDACC) and the Southern Hemisphere Additional Ozonesonde program (SHADOZ). To support users in their own verification of the fitness-for-purpose of the different datasets, a harmonized validation and reporting approach has been adopted, enabling comparative analysis of the validation results. Results for the Ozone_cci+ Level-2 and Level-3 limb profile CRDP are organized as follows

- Identification of the HARMOZ dataset, summarized in a table (Level-2 only);
- Study of the co-locations between HARMOZ and correlative datasets, and a brief discussion of the corresponding validation sample (Level-2 only);
- Presentation and analysis of the comparison results:
 - bias and spread as a function of altitude (or pressure), latitude and validation data source;
 - long-term stability over the entire time series as a function of altitude (or pressure);
 - dependence on other parameters, e.g. geophysical, instrument-related or auxiliary (Level-2 only);
 - summary table displaying the main quality indicators and other validation findings;
 - compliance with user requirements [RD8].

Summary plots in Section 6.4.19 give a comprehensive overview of uncertainties of all the limb ozone profile datasets of the Level-2 limb profile data portfolio, based on ground-based validation results. These plots shows systematic uncertainty, random uncertainty and drift estimates as a function of latitude and pressure, and their compliance with user requirements [RD8].



6.3 Conversion of profile representations

The native representation of the ozone profile data reported for satellite, ozonesondes, lidars and MWRs differs from each other (see Table 6.1). In order to minimize the number of manipulations of the Ozone_cci+ satellite data records under investigation, the validation analyses are performed –whenever possible– in the reported representation of the satellite ozone data record under study: ozone mole concentration (equivalent to ozone number density) at either fixed altitude (e.g. HARMOZ ALT) or fixed pressure levels (e.g. HARMOZ PRS). Conversions of the vertical or the ozone coordinate of satellite profile data is done using auxiliary meteorological information reported in these data files. It should be noted that the number density vs. altitude/pressure representation is not necessarily the native representation of the actual satellite ozone retrievals (as documented in the dataset identification table). In this case, the conversion of the satellite profile data was done by the retrieval teams.

Conversions of ozone partial pressure measurements versus pressure or altitude by ozonesonde into other representations is done using pressure, temperature and (in recent years) GPS altitude measured by the coupled radiosonde. Conversions of lidar ozone number density measurements versus altitude into other representations was done using ancillary pressure and/or temperature information extracted from ERA5 meteorological reanalyses provided by ECMWF. Observations of atmospheric temperature by different ground-based and satellite instruments are quite consistent in the troposphere and lower stratosphere, but they tend to diverge above ~30 km [RD78]. Reanalysis temperature data in middle and upper stratosphere are hence more uncertain, which constitutes a potential non-negligible source of (time-dependent) uncertainty in representation-converted ozone profile data. MWR typically retrieve ozone profiles as volume mixing ratio versus pressure, but number density, altitude and temperature are reported in the data file as well. Several sources of ancillary meteorological data are used throughout the MWR network (NCEP, ECMWF operational, ...) which may introduce uncertainty in the converted data sets. For this reason, all MWR validation analyses are performed in the pressure domain.

Table 6.1 summarizes the characteristics of each measurement source. Since the conversion of lidar data uses the same source of auxiliary data (ECMWF reanalysis) as that used for ESA and Third Party Mission data, the analyses remains insensitive to intrinsic deficiencies in the ERA5 data.

Table 6.1 - Characteristics of the satellite and ground-based measurements of the ozone profile. Bold indicates the native representation of the data files. In the case of HARMOZ data, this may differ from the native representation of the satellite ozone retrievals.

	Ozone unit	Altitude	Air pressure	Air temperature
HARMOZ PRS	Mole concentration	Retrieved / ERA-Interim / ERA5	Retrieved / ERA-Interim / ERA5	Retrieved / ERA-Interim / ERA5
HARMOZ ALT	Mole concentration	Retrieved / ERA-Interim / ERA5	Retrieved / ERA-Interim / ERA5	Retrieved / ERA-Interim / ERA5
Ozonesonde	Partial pressure	Provided	Measured	Measured
Stratospheric lidar	Number density	Measured	ERA5	ERA5
Microwave radiometer	Volume mixing ratio	Re/analysis	Retrieved	Re/analysis



6.4 *Level-2 limb profile products*

6.4.1 Validation method

The uncertainty of a single ozone profile by ground-based instruments is comparable to that obtained by satellite sensors. Nonetheless, a large sample of co-located satellite and ground-based profiles at numerous stations around the globe, allow us to derive meaningful estimates of systematic error, random uncertainty and long-term stability of the limb/occultation data records. The complementarity of the ozonesonde, lidar and MWR measurement techniques and network design allows further cross-checks which are crucial in achieving a robust assessment of satellite data quality.

As described in detail by Hubert et al. [RD43], the ground-based ozone profile observations are first screened according to the prescriptions by the data producers. Then, pairs of satellite and ground-based profiles that probe sufficiently overlapping air masses are searched for in a ± 6 h (MIPAS, Aura-MLS, SABER) or ± 12 h time window (all other instruments) and within 500 km radial distance (all instruments). The comparison to MWR profile data is done in a smaller window, less than 300 km and within 1-6h (depending on MWR station). When more than one satellite profile co-locates with a single ground-based profile, only the closest satellite profile is retained. Tightening the co-location window leaves the bias virtually unchanged and reduces the spread in the comparisons by a few percent. In a third step, the ozone unit is converted to mole concentration where necessary, see Section 6.3, after which sonde and lidar measurements are degraded to the vertical resolution of the satellite data set. To this end, a triangular window function is applied to smoothen each ground-based ozone profile in the altitude domain. The base width of the triangle is equal to the (altitude-dependent) vertical resolution reported along with the co-locating satellite profile. The shape of the window function was found to have a negligible impact on the final conclusions. This smoothing step is omitted for the MWR profiles since their resolution is poorer than that of the satellite data. Finally, the vertically smoothed ground-based profiles are regridded to the fixed altitude or pressure grid of HARMOZ, using the pseudo-inverse interpolation method by Calisesi et al. [RD33]. These pre-processing steps lead to a screened and co-located set of satellite and ground-based profiles in the same ozone coordinate and the same vertical coordinate and grid.

Statistical indicators for the analysis are derived as described in Sections 4.1 and 5.1 of Hubert et al. [RD43]. The selection of sites used for the drift analysis are those mentioned in this paper. The MWR drift analysis is considering only data from the Mauna Loa and Lauder site.



6.4.2 Envisat GOMOS ALGOM2s v1

Identification data record	
Observation principle	Stellar occultation, UV-visible spectral range
Platform	Envisat, mid-morning polar orbit, sun-synchronous precession
Responsible institute	Finnish Meteorological Institute – FMI
Contact person	Viktoria Sofieva (viktoria.sofieva@fmi.fi)
Coverage	
• Time	08/2002 – 12/2011
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 10 – 105 km (96 levels); PRS : 250 – 10 ⁻⁴ hPa (51 levels)
L1 processor and version	IPF v6
L2 processor and version	ALGOM2s v1
Validated L2 file version	ALT: fv0002* ; PRS : fv0001
Retrieval representation	O3 number density versus geometric altitude

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.2.1 Co-locations / Validation sample

Figure 6.1 shows the latitude–time distribution of the ozonesonde, lidar and microwave radiometer (MWR) measurements co-locating with GOMOS ALGOM2s v1 data. The sampling covers most latitude zones and is quite homogeneous in time, except for an interruption of GOMOS operations in the first half of 2005. Most GOMOS polar day profiles were screened out by the retrieval team at FMI, since twilight and bright limb viewing conditions degrade the quality of the retrieved profile.

6.4.2.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between GOMOS ALGOM2s v1 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.2 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.3 shows the same information calculated in 30° latitude zones.

The median bias is less than 4% between 20-45 km, except in the Arctic where it reaches -8 to -10% around 25 km and 28 km. Above the stratopause, GOMOS occultation and MWR data are within 6%. The meridian structure of the bias in the UTLS follows the meridian structure of the tropopause. Between the tropopause and 20 km the bias is positive in the Southern Hemisphere (5-10%) and negative at other latitudes (5-15%). At all latitudes a clear underestimation of ozone can be seen at the tropopause (-15% or more).

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above 20-25 km the half IP-68 spread is about 5% at the equator and increases gradually up to 10-15% at the poles. Below ~20 km the spread increases rapidly, reaching 40% around 18 km at the equator and around 11 km at the poles. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{ex-ante}$ (not shown here) provided in the GOMOS data files. The latter is 1-4% above ~20 km and increase rapidly up to 15-25% at lower levels.

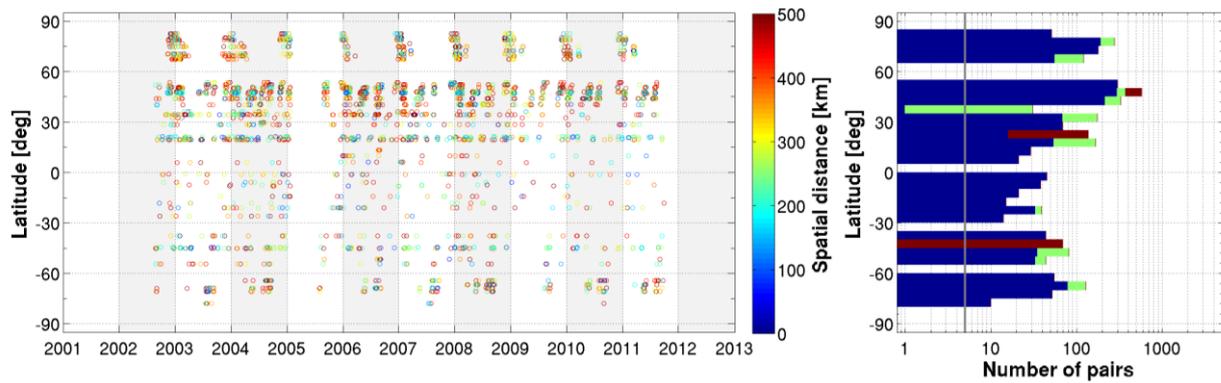


Figure 6.1 - (Left) Latitude–time distribution of co-locations between GOMOS ALGOM2s v1 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

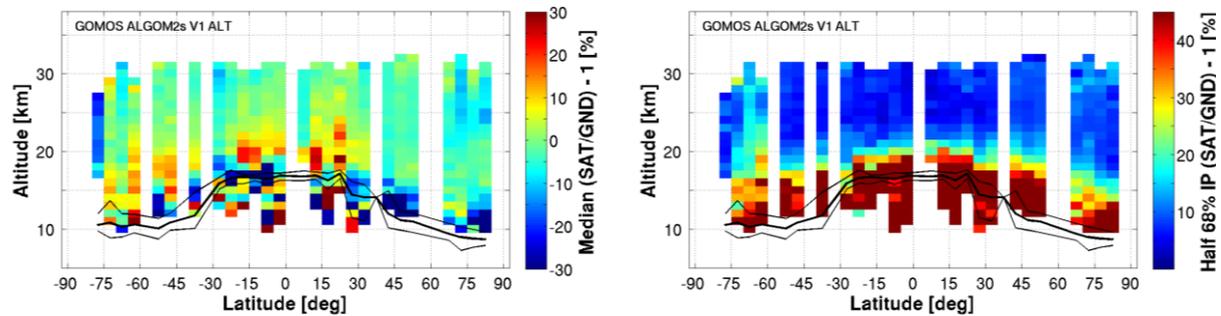


Figure 6.2 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between GOMOS ALGOM2s v1 ozone profile data and the global ozonesonde network, calculated over the entire GOMOS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

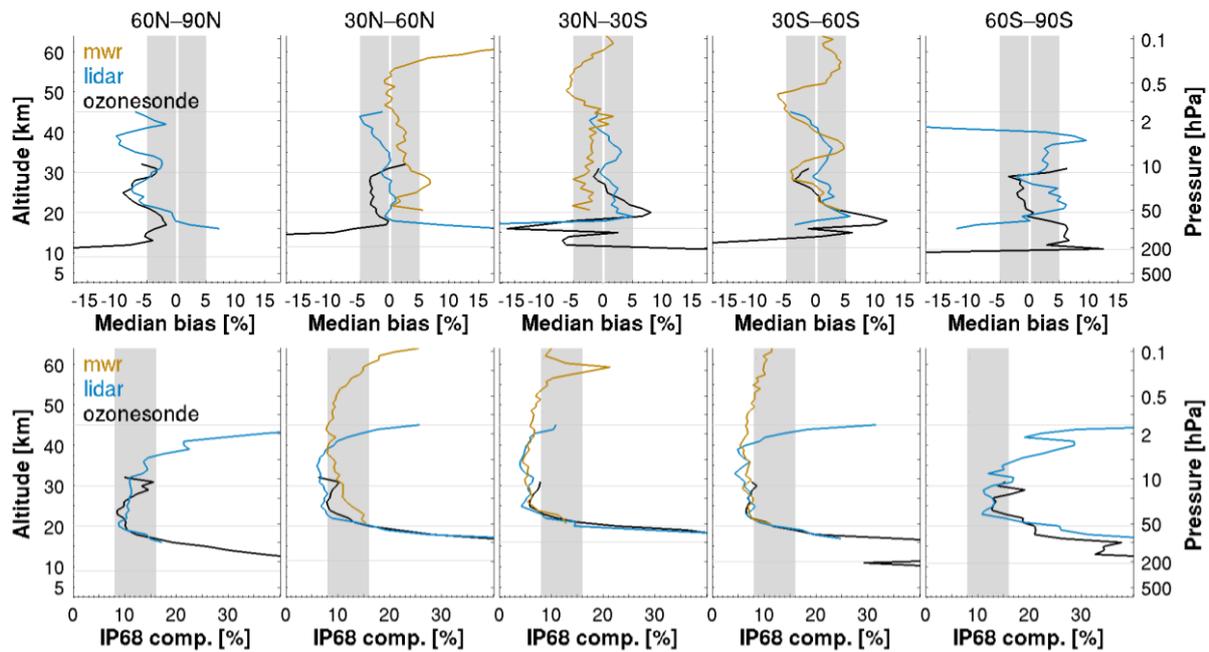


Figure 6.3 - (Top) Median bias between GOMOS ALGOM2s v1 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

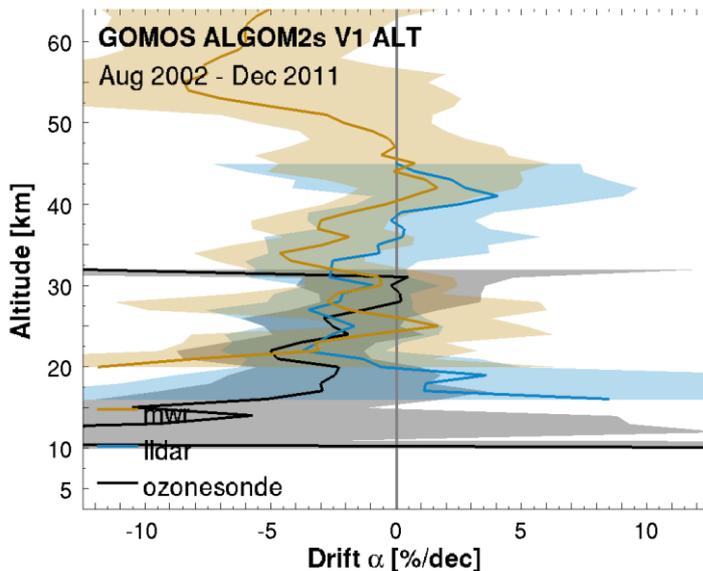


Figure 6.4 - Global average drift (in percent / decade) of GOMOS ALGOM2s v1 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2002-2011 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.2.3 Long-term stability

Figure 6.4 shows the drift of GOMOS ALGOM2s v1 ozone profile data with respect to co-located ozonesonde, lidar and MWR network data. Drift values are generally insignificant, negative and less than 3% per decade between ~25-50 km. The -8% per decade estimate around 55 km may be indicative of decreased stability in the mesosphere, though drift uncertainty is less reliable as only two MWR sites are used and uncertainties arise from the conversion of MWR VMR data to number density. Drift values in the lower stratosphere and UTLS region range from -5% to -10% per decade and reach the 2 σ threshold for the sonde comparisons



between 20-25 km. GOMOS drift relative to lower stratospheric lidar data, on the other hand, remains insignificant and flips sign at 20 km. Below ~20 km the uncertainty of the drift estimates increase rapidly due to increases in measurement noise and atmospheric variability; which explains the difference in sign between sonde and lidar results. This makes it difficult to obtain conclusive results in this part of the atmosphere.

6.4.2.4 Dependence of data quality on other parameters

The dependence of GOMOS ALGOM2s v1 data quality with respect to reference measurements has been studied also as a function of further parameters, including: star magnitude, star temperature, illumination condition, and occultation obliquity. There were no clear signs that the bias or comparison spread correlate with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

In addition, the bias, the comparison spread and the long-term stability of GOMOS ALGOM2s v1 ozone are in very good agreement in four different profile representations (combinations of altitude/pressure and number density/VMR). This indicates that the auxiliary pressure and temperature profiles included in the GOMOS HARMOZ data files and in the correlative data files are consistent.

6.4.2.5 Summary table of validation results

Table 6.2 - Summary of comparison results between GOMOS ALGOM2s v1 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-15, +7]	[-8, +5]	[-10, +10]	–
• Middle latitudes	[-20, +10]	[-3, +4]	[-5, +4]	[-5, +3]
• Tropics	[-20, +7]	[-2, +7]	[-3, +3]	[-6, +0]
Range comparison spread (half 68% interpercentile; %)				
• Polar	10 – >40	8–12/20	12–30	–
• Middle latitudes	10 – >45	7–10	6–9	6–15
• Tropics	10 – >45	5–10	4–6	5–14
Range long-term stability (drift; %/decade)				
• Ground network	[-12, +3]	[-4, 0]	[-3, +2]	[-8, 0]



6.4.2.6 Compliance with user requirements

From the results reported above it can be concluded that the GOMOS ALGOM2s v1 ozone profile data record is compliant with all but one sampling and resolution requirement. Data quality does not meet the requirements in the lowermost part of the ozone profile, but improves at higher altitudes. User requirements for random uncertainty are met for 60°N-60°S between 25-50 km and for decadal stability between 20-50 km. At lower and higher altitudes there are indications of a negative drift that is not compliant with requirements, though individual results are not significant. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over the entire vertical range, even though this may not reflect the actual data quality.

Table 6.3 - Compliance of GOMOS ALGOM2s v1 with user requirements (URD v3.1).

	User requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	order of 100-250 km [RD88]
Vertical resolution	< 1–3 km	2–3 km
Observation frequency	< 3 days	3 days
Time period	(1980-2010) – (2003-2010)	08/2002 – 12/2011
Total uncertainty in height registration	< ± 500 m	60–150 m
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
User requirement	< 8-16 %				< 8%				
	Uncertainty including only random component								
• Arctic					(large natural variability)				
• Mid NH									
• Tropics					(large natural variability)				
• Mid SH									
• Antarctic					(large natural variability)				
User requirement	Long-term stability								
	< 1-3% per decade								
• Ground network									



6.4.3 Envisat GOMOS Bright Limb v1.2

Identification data record	
Observation principle	UV-visible scattering at bright limb
Platform	Envisat, mid-morning polar orbit, sun-synchronous precession
Responsible institute	Finnish Meteorological Institute – FMI
Contact person	Simo Tukiainen (simo.tukiainen@fmi.fi)
Coverage	
• Time	04/2002 – 04/2012
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 19 – 59 km (41 levels); PRS : 70 – 0.1 hPa (21 levels)
L1 processor and version	IPF v6
L2 processor and version	GBL v1.2
Validated L2 file version	ALT: fv0001*; PRS : fv0001
Retrieval representation	O3 number density versus geometric altitude

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.3.1 Co-locations / Validation sample

Figure 6.5 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with GOMOS Bright Limb v1.2 data. The sampling covers most latitude zones and is quite homogeneous in time, except a short interruption of GOMOS operations in the first half of 2005. It is well distributed over the time series and over seasons, except at high latitudes during polar night.

6.4.3.2 Bias and spread

The median bias and half the 68% interpercentile range of the relative difference between GOMOS BL v1.2 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.6 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.7 show the same information calculated in 30° latitude zones.

GOMOS bright limb ozone data is biased low relative to ground-based data between 25-30 km and the stratopause. The underestimation varies slightly with latitude and becomes clearly larger with increasing altitude until ~40 km, where it reaches -10% to -15%. A rapid transition to positive bias values is seen towards the stratopause and extends well into the mesosphere. Below 25-30 km, there is an overestimation of up to 5% in the Northern Hemisphere and the tropics, but not at more southern latitudes. Our analysis corroborates the findings by [RD86]. We refer to latter publication for a discussion of the origin of the bias structure.

The structure and magnitude of the comparison spread is very similar to that of the ALGOM2s data. Below 25-30 km the bright limb data are slightly more noisy than GOMOS occultation data, the comparison spread is ~1-2% larger. The ex-ante uncertainty (not shown here) for both GOMOS data sets is generally very similar, except between 32-40 km and above the stratopause where the expected uncertainty is clearly larger (by ~5-10%).

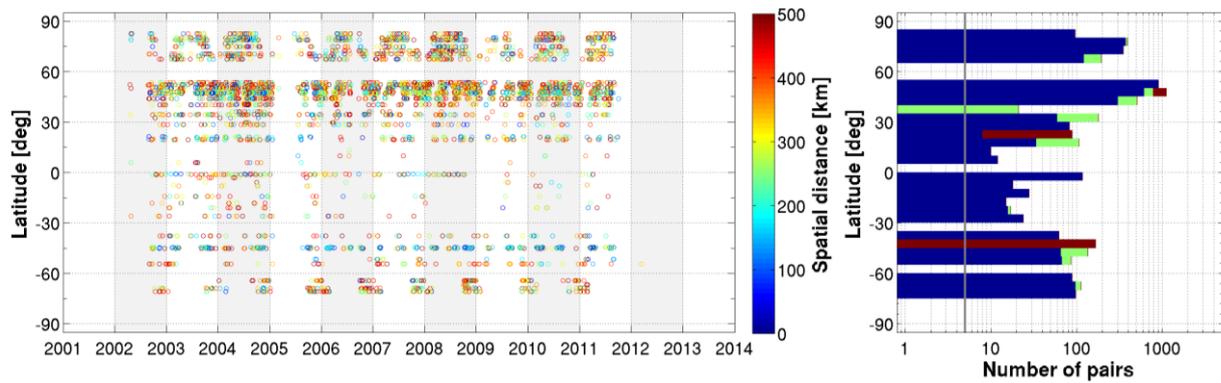


Figure 6.5 - (Left) Latitude–time distribution of co-locations between GOMOS Bright Limb v1.2 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

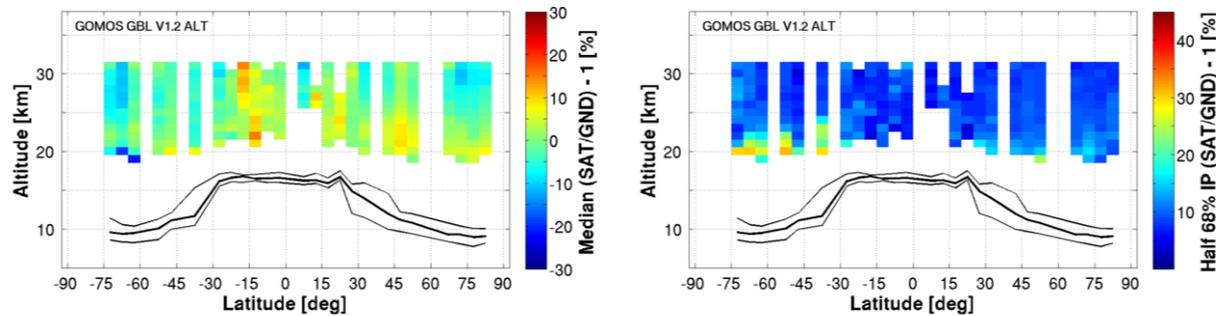


Figure 6.6 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between GOMOS Bright Limb v1.2 ozone profile data and the global ozonesonde network, calculated over the entire GOMOS time period and in 5° bins. Black lines indicate the median (thick) and 1 σ spread (thin) of the tropopause altitude.

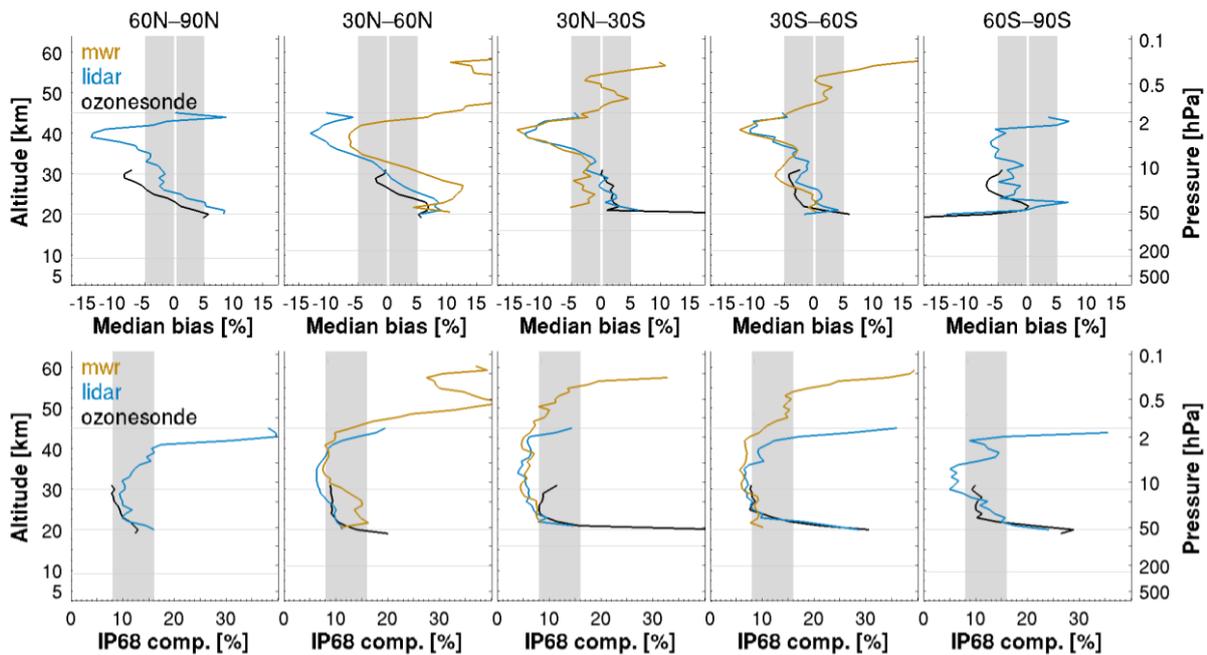


Figure 6.7 - (Top) Median bias between GOMOS Bright Limb v1.2 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

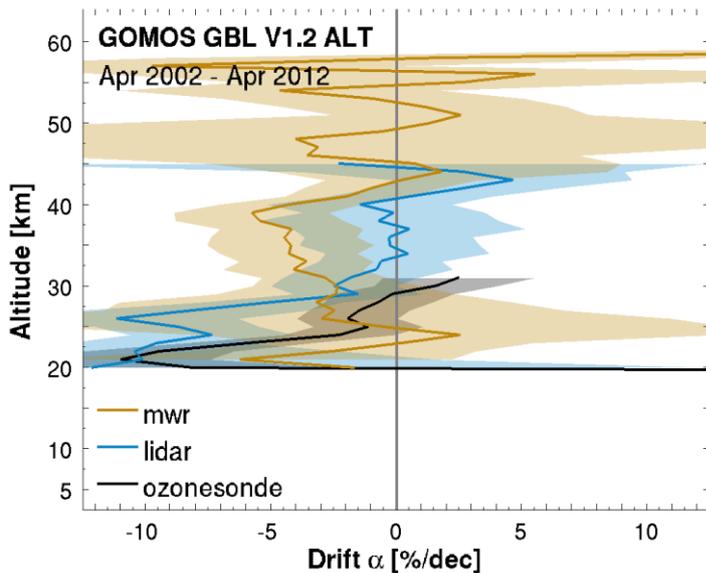


Figure 6.8 - Global average drift (in percent / decade) of GOMOS Bright Limb v1.2 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2002-2012 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.3.3 Long-term stability

Figure 6.8 shows the drift of GOMOS Bright Limb v1.2 ozone profile data with respect to the co-located ground-based data. The sign of the drift is generally negative between 20-40 km, but quantitative results and significance vary by reference instrument. The most compelling evidence of negative drift is found below ~25 km where both lidar and sonde results are significant. Drift is about -10% per decade at 21 km and about -1 to -3% per decade at 30 km. MWR comparisons over this vertical range are more noisy but the derived drift is broadly consistent with the other results. The stability between 30-40 km is more difficult to assess



given the tension between lidar and MWR results. The former is based on more than a handful of stations and does not show a significant drift. The latter, on the other hand, suggest a significant drift of -4% per decade. But since that is based on just two sites we are inclined to put more confidence in the lidar results. In the mesosphere no drift is noted either.

6.4.3.4 Dependence of data quality on other parameters

The bias, the comparison spread and the long-term stability of GOMOS BL v1.2 ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the GOMOS HARMOZ data files and in the correlative data files are consistent.

6.4.3.5 Summary table of validation results

Table 6.4 - Summary of comparison results between GOMOS Bright Limb v1.2 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	–	[-8, +8]	[-14, +9]	–
• Middle latitudes	–	[-4, +8]	[-13, +6]	[-5, +20]
• Tropics	–	[-1, +4]	[-13, 0]	[-3, +10]
Range comparison spread (half 68% interpercentile; %)				
• Polar	–	8-15/25	6-35	–
• Middle latitudes	–	8-15	7-12	12-30
• Tropics	–	6-15	5-7	7-30
Range long-term stability (drift; %/decade)				
• Ground network	–	[-10, +1]	[-5, +5]	[-9, +7]



6.4.3.6 Compliance with user requirements

From the results reported above it can be concluded that the GOMOS Bright Limb v1.2 ozone profile data record is compliant with most of sampling and resolution requirements. No data are provided in the lower stratosphere, below 20 km. Requirements on random uncertainty are met between 25-40 km and partially slightly above and below. Requirements on decadal stability are met between 30-45 km, at lower altitudes a large negative drift is observed which becomes significant below 25 km. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over the entire vertical range in the Arctic, even though this may not reflect the actual data quality.

Table 6.5 - Compliance of GOMOS Bright Limb v1.2 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	300–400 km [RD87]
Vertical resolution	< 1–3 km	2–3 km [RD86]
Observation frequency	< 3 days	3 days
Time period	(1980–2010) – (2003–2010)	04/2002 – 04/2012
Total uncertainty in height registration	< ± 500 m	60–150 m
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %		< 8%						
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 			(large natural variability)						
Long-term stability									
User requirement	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.4 Envisat MIPAS IMK/IAA v7

The Ozone_cci+ Level-2 CRDP includes MIPAS nominal mode observations acquired at full spectral resolution between July 2002 and March 2004 (L2 processor version V7H_O3_40) and at reduced resolution between January 2005 and April 2012 (L2 processor version V7R_O3_240). There is an (altitude-dependent) bias between MIPAS O3 retrievals in both periods, which is why the validation analysis is differentiated each period. Below, we will refer to both MIPAS data records as version 7.

Results for an updated MIPAS data record (v8) are reported in Section 6.4.5.

Identification data record	
Observation principle	Infrared limb emission, along track scanning
Platform	Envisat, mid-morning polar orbit, sun-synchronous precession
Responsible institute	Karlsruhe Institute of Technology - KIT
Contact person	Alexandra Laeng (alexandra.laeng@kit.edu)
Coverage	
• Time	07/2002 – 03/2004 (Full Resolution period), 01/2005 – 04/2012 (Reduced Resolution period)
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 70 km (65 levels); PRS : 500 – 0.04 hPa (39 levels)
L1 processor and version	ESA IPF v7.11
L2 processor and version	IMK-IAA V7H_O3_40 (2002-2004) and V7R_O3_240 (2005-2012)
Validated L2 file version	ALT : fv0002*; PRS : fv0002
Retrieval representation	O3 volume mixing ratio versus geometric altitude

*All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.4.1 Co-locations / Validation sample

Figure 6.9 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with MIPAS nominal mode measurements for the full spectral resolution period (2002-2004) and the reduced spectral resolution period (2005-2012). The sampling covers most latitude bands and is homogeneous in time, except for an interruption of operations in 2004 and the reduced duty cycle during 2005-2006.

6.4.4.2 Bias and spread

The median bias and half the 68% interpercentile range of the relative difference between MIPAS IMK-IAA v7 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.10 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.11 and Figure 6.12 show the same information calculated in 30° latitude zones for both periods.

MIPAS bias differs considerably for both periods with a sign and magnitude that is altitude-dependent. Compared to the 2005-2012 data, ozone values during 2002-2004 are up to 5% larger below 20-25 km and up to 5% smaller at higher altitudes. A positive bias of 3-7% relative to ground-based measurements is seen for both periods in the middle and upper stratosphere. It typically transitions to negative values in the lowermost stratosphere and close to the stratopause. Below the tropopause vertical oscillations (10% peak to peak) are noted in the bias profile relative to ozonesonde.

The meridian structure of the comparison spread $s_{\Delta x}$ is similar for both periods and follows at first order the meridian structure of the tropopause. In the middle and upper stratosphere the spread is about 4% at the equator, and increases to 8-10% at the poles. Comparison spread increases rapidly below 20 km, reaching



40% right above the tropopause. The spread seen in the comparisons is several percent larger than the ex-ante random uncertainty $S_{ex-ante}$ provided in the MIPAS data files (not shown here). The latter is 2-8% in the mesosphere, 1-2% in the stratosphere and rapidly increases below 20 km altitude to 20-30% in the upper troposphere.

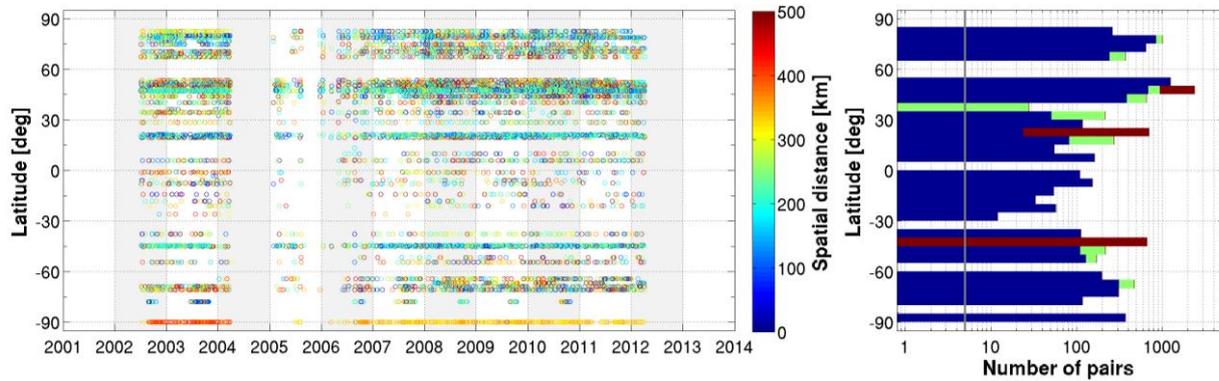


Figure 6.9 - (Left) Latitude–time distribution of co-locations between MIPAS IMK-IAA v7 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC microwave radiometer). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

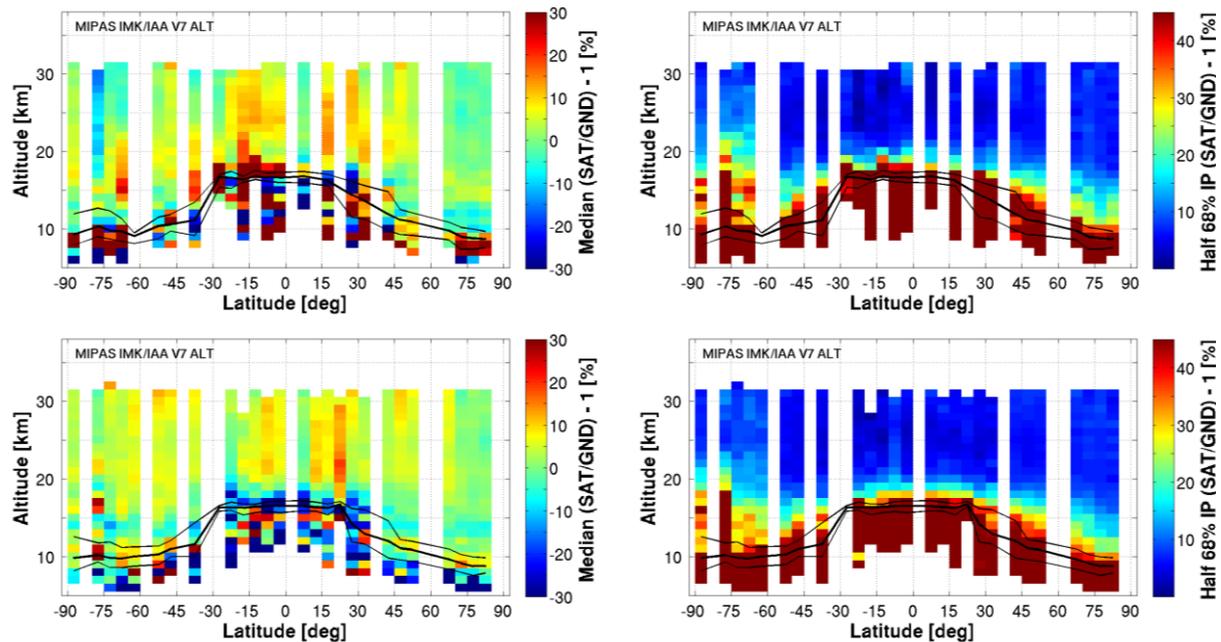


Figure 6.10 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between MIPAS IMK-IAA v7 ozone profile data and the global ozonesonde network, calculated over the 2002-2004 (top row) and 2005-2012 (bottom) time periods and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

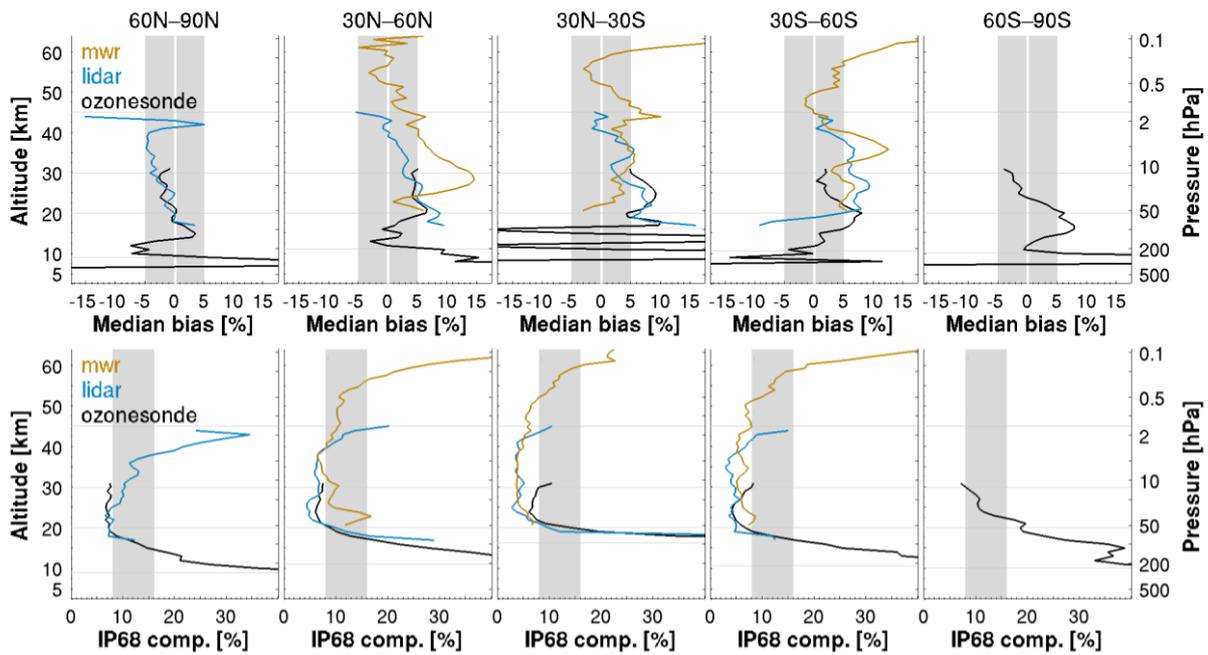


Figure 6.11 - (Top) Median bias between MIPAS IMK-IAA v7 (2002-2004) ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

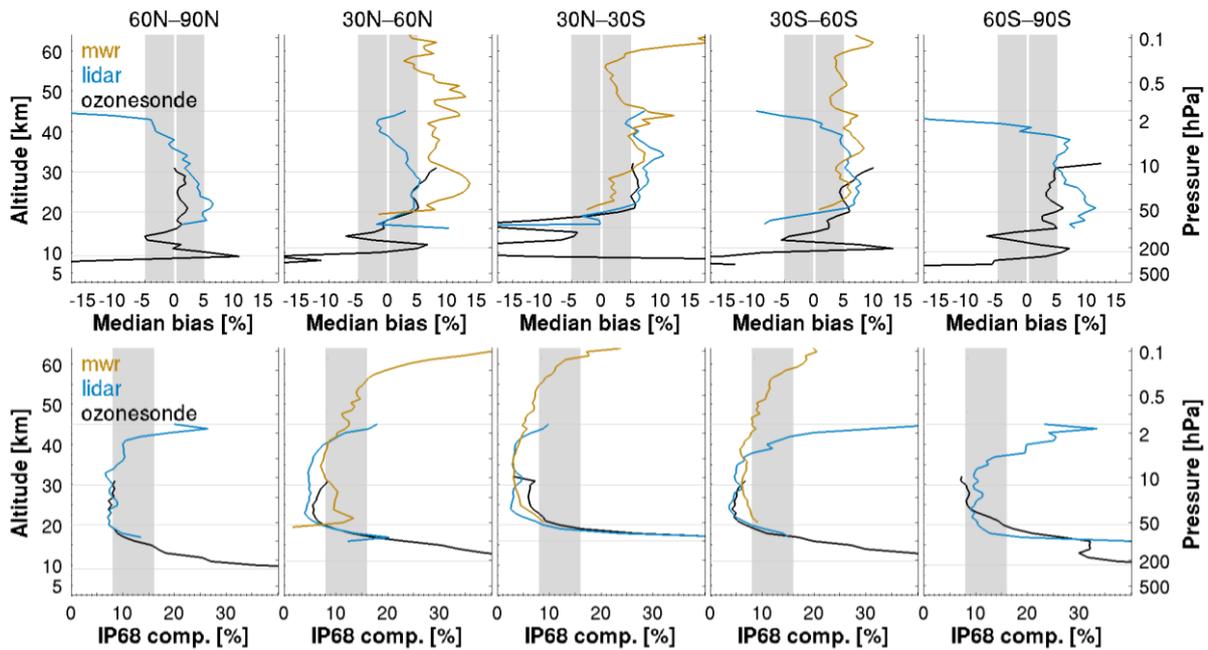


Figure 6.12 - As Figure 6.11, but for MIPAS data during the 2005-2012 period.

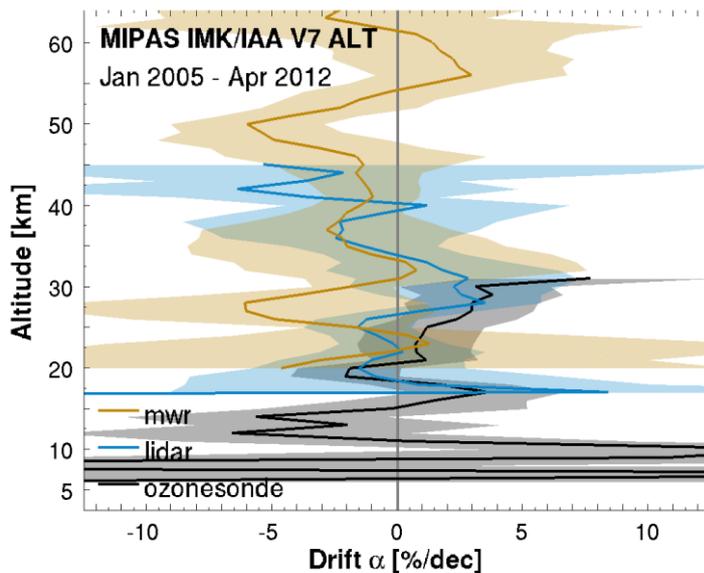


Figure 6.13 - Global average drift (in percent / decade) of MIPAS IMK-IAA v7 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated for the 2005-2012 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.4.3 Long-term stability

Figure 6.13 shows the drift estimates of MIPAS IMK-IAA v7 ozone profile data with respect to the co-located ozonesonde, lidar and MWR network data [RD56]. The 2002-2004 period is not taken into account since it is too short for trend studies and, more importantly, the bias between the periods introduces artefacts when not accounted for in the regression [RD36]. Drift estimates between 15-40 km are insignificant and less than $\pm 2\%$ per decade. Below 15 km, results from comparison to ozonesonde are statistically consistent with a no-drift hypothesis. Comparison to lidar and microwave radiometers, on the other hand, hint at a negative MIPAS drift of 3-4% per decade in the uppermost stratosphere and lower mesosphere. However, above 40 km altitude, the drift estimates are, either, not well constrained due to considerable station-to-station variability (lidar), or, they are possibly overestimated as a result of too few stations (MWR). We advise caution to MIPAS users who require stable measurements above 40 km.

6.4.4.4 Dependence of data quality on other parameters

Further parameters which possibly impact the MIPAS data quality were studied, including the retrieval output parameters χ^2 (normalized chi-square of retrievals), rms (root mean square of residual spectra) and DoF (degrees of freedom of target retrieval). There were no clear signs that the bias or comparison spread correlates with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of MIPAS ozone are very similar in four different profile representations (not shown here). This indicates that the auxiliary pressure and temperature profiles included in the MIPAS HARMOZ data files and in the correlative data files are consistent.



6.4.4.5 Summary table of validation results (2005-2012 period)

Table 6.6 - Summary of comparison results between MIPAS IMK-IAA v7 (2005-2012 period) and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-5, +8]	[0, +6]	[-10, +6]	–
• Middle latitudes	[-7, +7]	[+4, +7]	[-2, +7]	[+4, +10]
• Tropics	[-20, 0]	[+1, +7]	[+6, +8]	[+1, +7]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8 – >40	7–12	8–27	–
• Middle latitudes	8 – >45	5–8	5–12	8–24
• Tropics	10 – >45	3–8	3–5	5–13
Range long-term stability (drift; %/decade)				
• Ground network	[-4, +1]	[-1, +2]	[-3, +1]	[-6, +3]

6.4.4.6 Compliance with user requirements (2005-2012 period)

From the results reported above we conclude that the MIPAS IMK-IAA v7 limb ozone profile data record is (nearly) compliant with most of sampling and resolution requirements. Data quality does not meet the requirements in the lowermost part of the ozone profile, below 15 km. The random uncertainty meets the 8% or 8-16% requirement between 20 and 40-50 km. The stability of the MIPAS data record over 2005-2012 complies with the requirement over most of the profile. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over the part of the vertical range, even though this may not reflect the actual data quality.

Table 6.7 - Compliance of MIPAS IMK-IAA v7 (2005-2012 period) with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	200–400 km [RD90]
Vertical resolution	< 1–3 km	2.5–5 km [RD55]
Observation frequency	< 3 days	3 days
Time period	(1980-2010) – (2003-2010)	07/2002 – 04/2012
Total uncertainty in height attribution	< ± 500 m	[RD50]
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic									
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic									
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.5 Envisat MIPAS IMK/IAA v8

The Ozone_cci+ Level-2 CRDP includes MIPAS nominal mode observations acquired at full spectral resolution between July 2002 and March 2004 (L2 processor version V8H_O3_61) and at reduced spectral resolution between January 2005 and April 2012 (L2 processor version V8R_O3_261). There is an (altitude-dependent) bias between MIPAS O3 retrievals in both periods, which is why the validation analysis is differentiated each period. Below, we will refer to both MIPAS data records as version 8.

Assessment of earlier versions of the MIPAS IMK/IAA data set is reported in the preceding Section 6.4.4 (v7).

Identification data record	
Observation principle	Infrared limb emission, along track scanning
Platform	Envisat, mid-morning polar orbit, sun-synchronous precession
Responsible institute	Karlsruhe Institute of Technology - KIT
Contact person	Alexandra Laeng (alexandra.laeng@kit.edu)
Coverage	
• Time	07/2002 – 03/2004 (Full Resolution period), 01/2005 – 04/2012 (Reduced Resolution period)
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 70 km (65 levels); PRS : 500 – 0.04 hPa (39 levels)
L1 processor and version	ESA MICAL v8.03
L2 processor and version	IMK-IAA V8H_O3_61 (2002-2004) and V8R_O3_261 (2005-2012)
Validated L2 file version	ALT : fv0001
Retrieval representation	O3 volume mixing ratio versus geometric altitude

6.4.5.1 Co-locations / Validation sample

Figure 6.14 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with MIPAS nominal mode measurements for the full spectral resolution period (2002-2004) and the reduced spectral resolution period (2005-2012). The sampling covers most latitude bands and is homogeneous in time, except for an interruption of operations in 2004 and the reduced duty cycle during 2005-2006.

6.4.5.2 Bias and spread

The median bias and half the 68% interpercentile range of the relative difference between MIPAS IMK-IAA v8 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.15 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.16 and Figure 6.17 show the same information calculated in 30° latitude zones for both periods.

MIPAS bias differs considerably for both periods with a sign and magnitude that is altitude-dependent. Compared to the 2005-2012 data, ozone values during 2002-2004 are up to 5% larger below 20-25 km and up to 5% smaller at higher altitudes. A positive bias of 4-8% relative to ground-based measurements is seen for both periods in the middle and upper stratosphere. It typically transitions to negative values in the lowermost stratosphere and around the stratopause. Below the tropopause vertical oscillations (10% and more, peak to peak) are noted in the bias profile relative to ozonesonde.

The meridian structure of the comparison spread $s_{\Delta x}$ is similar for both periods and follows at first order the meridian structure of the tropopause. In the middle and upper stratosphere the spread is about 4% at the



equator, and increases to 8-10% at the poles. Comparison spread increases rapidly below 20 km, reaching 40% right above the tropopause. The spread seen in the comparisons is several percent larger than the ex-ante random uncertainty $S_{\text{ex-ante}}$ provided in the MIPAS data files (not shown here). The latter is 2-8% in the mesosphere, 1-2% in the stratosphere and rapidly increases below 20 km altitude to 20-30% in the upper troposphere.

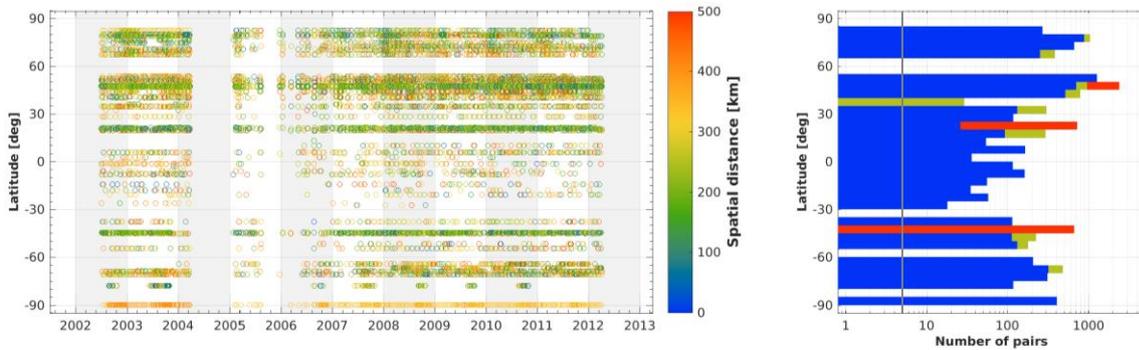


Figure 6.14 - (Left) Latitude–time distribution of co-locations between MIPAS IMK-IAA v8 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC microwave radiometer). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

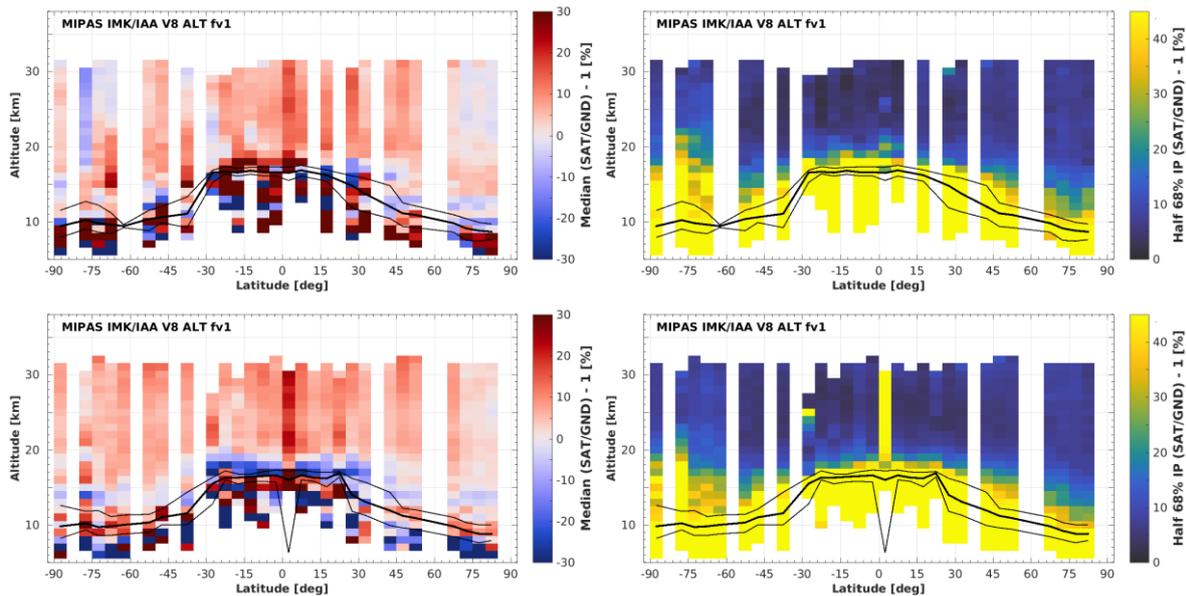


Figure 6.15 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between MIPAS IMK-IAA v8 ozone profile data and the global ozonesonde network, calculated over the 2002-2004 (top row) and 2005-2012 (bottom) time periods and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

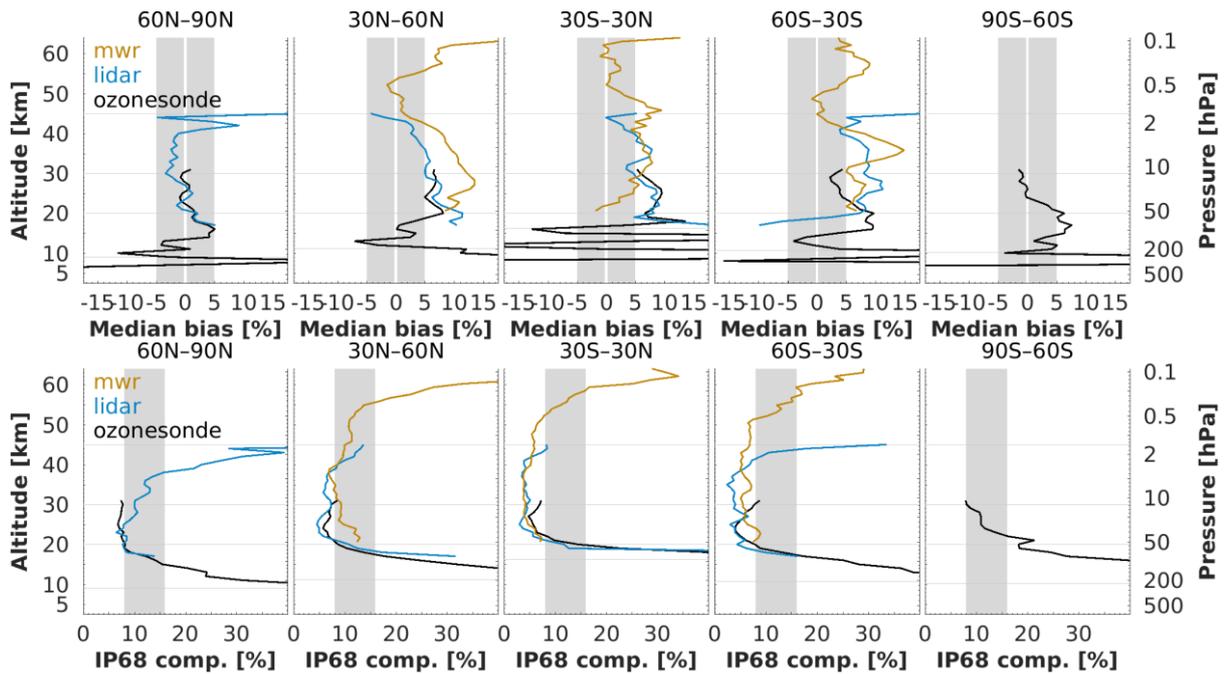


Figure 6.16 - (Top) Median bias between MIPAS IMK-IAA v8 (2002-2004) ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

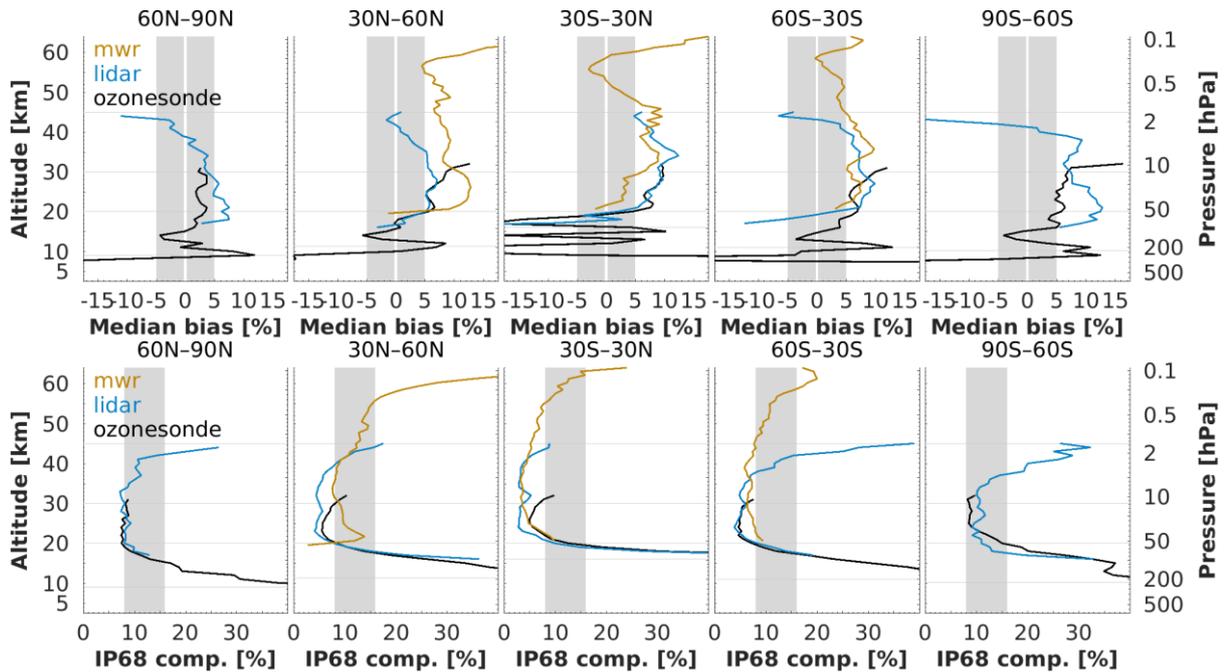


Figure 6.17 - As Figure 6.16, but for MIPAS data during the 2005-2012 period.

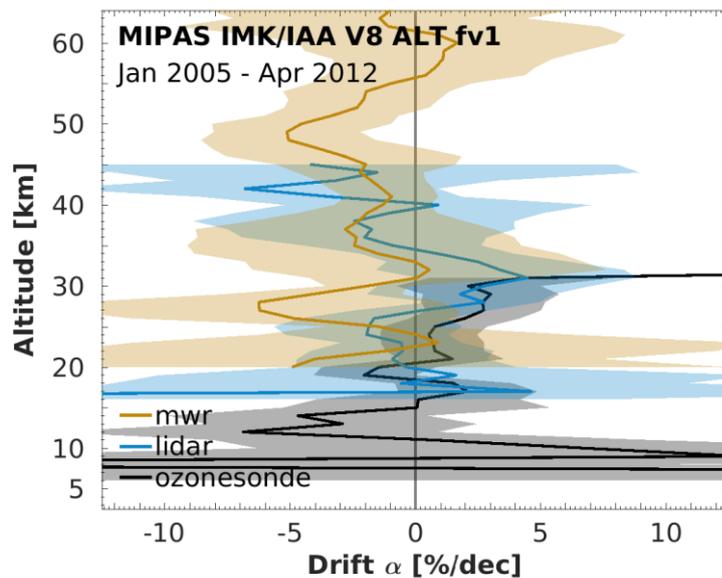


Figure 6.18 - Global average drift (in percent / decade) of MIPAS IMK-IAA v8 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated for the 2005-2012 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.5.3 Long-term stability

Figure 6.18 shows the drift estimates of MIPAS IMK-IAA v8 ozone profile data with respect to the co-located ozonesonde, lidar and MWR network data. The 2002-2004 period is not taken into account since it is too short for trend studies and, more importantly, the bias between the periods introduces artefacts when not accounted for in the regression [RD36]. Drift estimates between 15-40 km are insignificant and less than $\pm 2\%$ per decade. Below 15 km, results from comparison to ozonesonde are statistically consistent with a no-drift hypothesis. Comparison to lidar and microwave radiometers, on the other hand, hint at a negative MIPAS drift of 3-4% per decade in the uppermost stratosphere and lower mesosphere. However, above 40 km altitude, the drift estimates are, either, not well constrained due to considerable station-to-station variability (lidar), or, they are possibly overestimated as a result of too few stations (MWR). We advise caution to MIPAS users who require stable measurements above 40 km.

6.4.5.4 Dependence of data quality on other parameters

Further parameters which possibly impact the MIPAS data quality were studied, including the retrieval output parameters χ^2 (normalized chi-square of retrievals), rms (root mean square of residual spectra) and DoF (degrees of freedom of target retrieval). There were no clear signs that the bias or comparison spread correlates with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of MIPAS ozone are very similar in four different profile representations (not shown here). This indicates that the auxiliary pressure and temperature profiles included in the MIPAS HARMOZ data files and in the correlative data files are consistent.



6.4.5.5 Summary table of validation results (2005-2012 period)

Table 6.8 - Summary of comparison results between MIPAS IMK-IAA v8 (2005-2012 period) and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-6, +7]	[+2, +9]	[-10, +6]	–
• Middle latitudes	[-7, +7]	[+4, +7]	[-2, +7]	[+4, +10]
• Tropics	[-20, 0]	[+3, +7]	[+6, +8]	[+1, +9]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8 – >40	8–12	8–27	–
• Middle latitudes	8 – >45	5–8	5–12	8–25
• Tropics	10 – >45	3–8	4–6	6–13
Range long-term stability (drift; %/decade)				
• Ground network	[-5, 0]	[-1, +1]	[-3, +2]	[-5, +2]

6.4.5.6 Compliance with user requirements (2005-2012 period)

From the results reported above it can be concluded that the MIPAS IMK-IAA v8 limb ozone profile data record is (nearly) compliant with most of sampling and resolution requirements. Data quality does not meet the requirements in the lowermost part of the ozone profile, below 15 km. The random uncertainty meets the 8% or 8-16% requirement between 20 and 40-50. The stability of the MIPAS data record over 2005-2012 complies with the requirement over most of the profile. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over the part of the vertical range, even though this may not reflect the actual data quality.

Table 6.9 - Compliance of MIPAS IMK-IAA v8 (2005-2012 period) with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	200–400 km [RD90]
Vertical resolution	< 1–3 km	2.5–5 km [RD55]
Observation frequency	< 3 days	3 days
Time period	(1980-2010) – (2003-2010)	07/2002 – 04/2012
Total uncertainty in height attribution	< ± 500 m	[RD50]
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic									
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic									
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.6 Envisat SCIAMACHY UBr v3.5

Identification data record	
Observation principle	UV-visible limb scattering, along track scanning
Platform	Envisat, mid-morning polar orbit, sun-synchronous precession
Responsible institute	University of Bremen - IUP
Contact person	Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	08/2002 – 04/2012
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 5 – 65 km (61 levels); PRS : 450 – 0.05 hPa (36 levels)
L1 processor and version	ESA SGP v8.01
L2 processor and version	UBr v3.5
Validated L2 file version	ALT : fv0005*; PRS : fv0004
Retrieval representation	O3 number density versus geometric altitude

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.6.1 Co-locations / Validation sample

Figure 6.19 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with the SCIAMACHY UBr v3.5 dataset. The sampling is very dense and covers all but the highest latitude zones. It is well distributed over the entire time series and over seasons, except poleward of 70° latitude where SCIAMACHY cannot observe the dark limb during polar night.

6.4.6.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between SCIAMACHY UBR v3.5 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.20 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.21 show the same information calculated in 30° latitude zones.

Sonde and lidar comparisons show a very coherent picture of SCIAMACHY bias. It is generally less than 5% between 20–30 km, positive southward of 30°S and positive elsewhere. Above 30 km, all lidar and MWR results indicate an increasing positive bias with altitude which peaks at 10–15% around the stratopause. In the mesosphere, the slope of the bias profile changes sign at all four MWR sites considered. In the UTLS and below, a negative bias of ~10% is found in the tropics and Arctic; at other latitudes a 10% overestimation is noted.

The meridian structure of the spread $s_{\Delta x}$ in comparisons of SCIAMACHY to ground-based data is quite common. In the middle and upper stratosphere the spread is generally about 5% at the equator, but increases gradually to 10% at the poles. In the lower stratosphere the comparison spread increases rapidly, reaching 40% around 16 km at the equator and around 9 km at higher latitudes. The ex-ante random uncertainties $s_{ex-ante}$ provided in the SCIAMACHY record (not shown here) are 2–3% above 20 km.

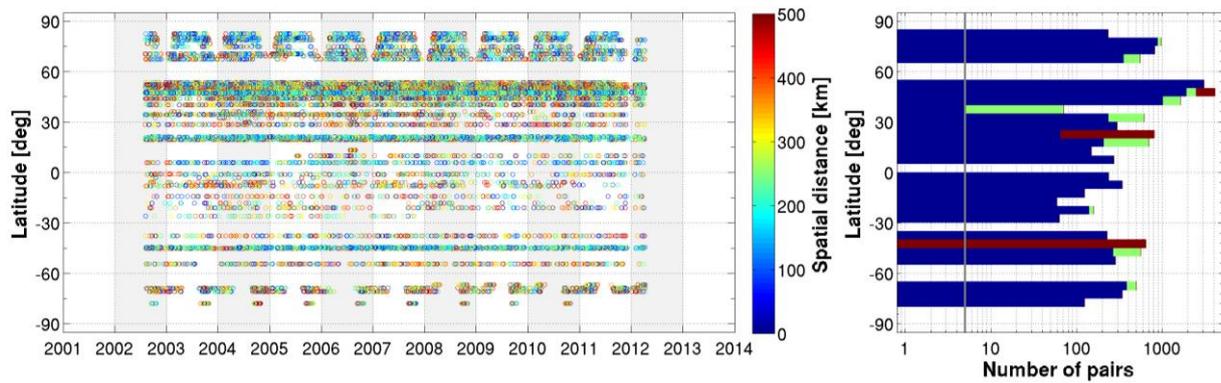


Figure 6.19 - (Left) Latitude–time distribution of co-locations between SCIAMACHY UBr v3.5 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

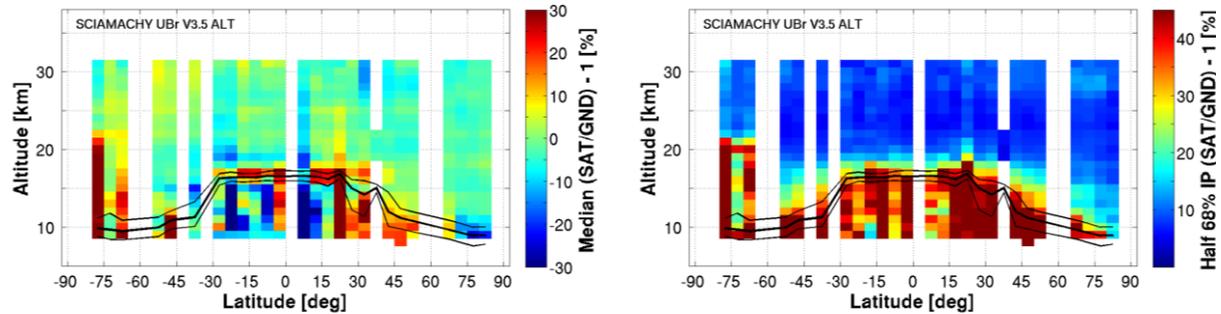


Figure 6.20 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between SCIAMACHY UBr v3.5 ozone profile data and the global ozonesonde network, calculated over the entire SCIAMACH time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

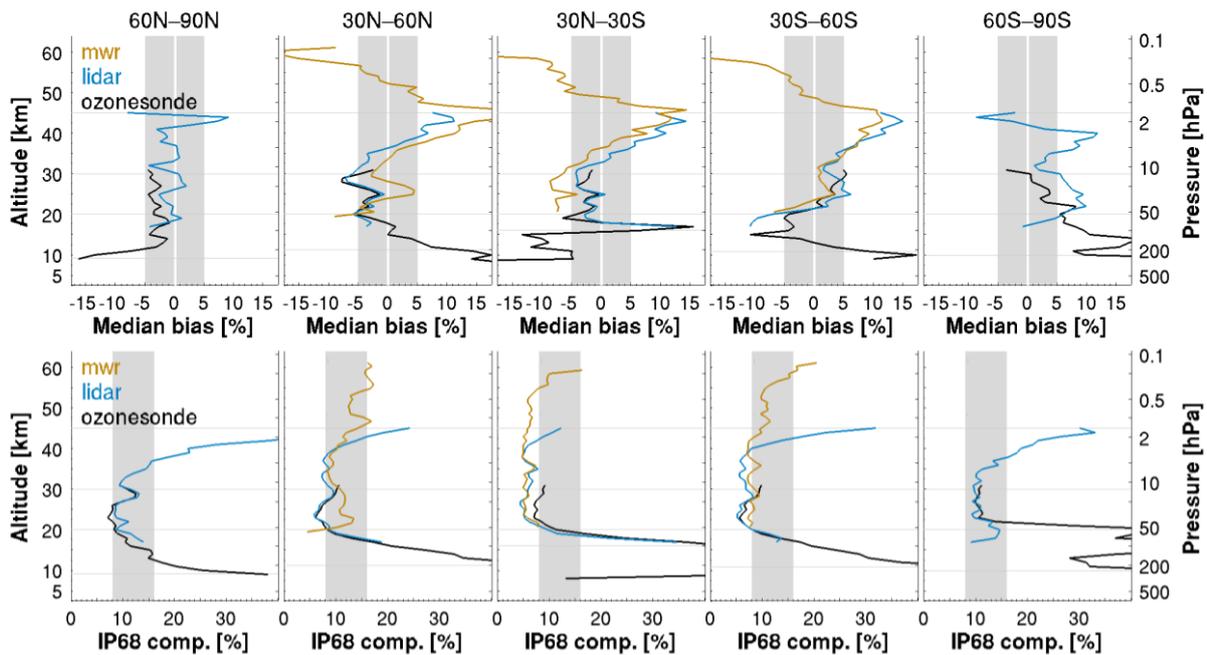


Figure 6.21 - (Top) Median bias between SCIAMACHY UBr v3.5 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

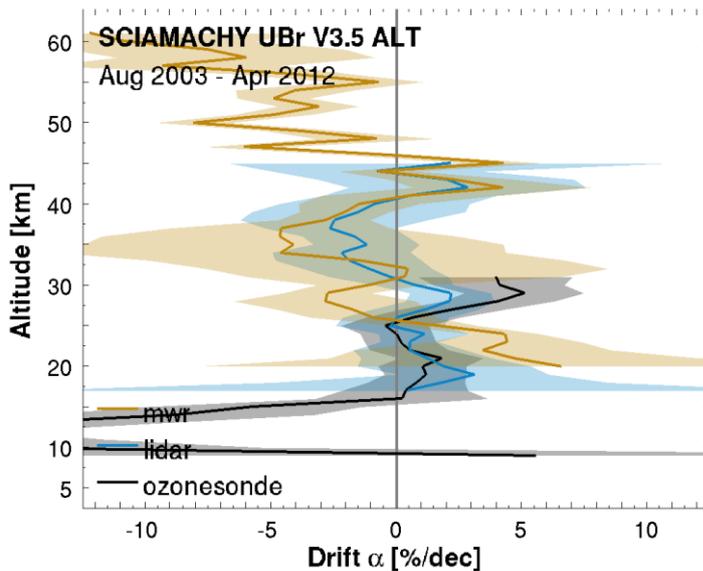


Figure 6.22 - Global average drift (in percent / decade) of SCIAMACHY UBr v3.5 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2003-2012 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.6.3 Long-term stability

Pointing problems in the first year of SCIAMACHY operations may have led to reduced stability of the ozone profile data record. Here, we follow the suggestion by Sofieva et al. (2017, [RD81]) to exclude measurements before August 2003 from the analysis. Figure 6.22 shows that the stability of the screened SCIAMACHY UBr v3.5 ozone profile data record (Aug 2003 – Apr 2012) is better than 3% per decade over a large part of the stratosphere. There are indications of negative drift of more than 5% per decade below 15 km, above 45 km and perhaps around 35 km. However, significance may be either underestimated (mesosphere, only two MWR sites) or insufficient but with good agreement for both reference records (middle stratosphere). The



most convincing sign of instability is in the UTLS, with a negative drift of more than 10% per decade between 10-15 km.

6.4.6.4 Dependence of data quality on other parameters

Further parameters which possibly impact the SCIAMACHY data quality were studied, including solar zenith angle. There were no clear signs that the bias or comparison spread correlate with this parameter. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of SCIAMACHY ozone are very similar in four different profile representations (not shown here). This indicates that the auxiliary pressure and temperature profiles included in the SCIAMACHY HARMOZ data files and in the correlative data files are consistent.

6.4.6.5 Summary table of validation results

Table 6.10 - Summary of comparison results between SCIAMACHY UBr v3.5 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-15, +15]	[-4, +6]	[-3, +7]	–
• Middle latitudes	[-10, +10]	[-8, +4]	[-5, +12]	[-15, +10]
• Tropics	[-10, +15]	[-4, 0]	[-5, +12]	[-10, +12]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8 – >40	7–12	10–33	–
• Middle latitudes	8 – >40	5–9	6–13	10–16
• Tropics	9 – >50	5–9	5–7	6–10
Range long-term stability (drift; %/decade)				
• Ground network	[-12, +3]	[0, +2]	[-4, +2]	[-9, +2]

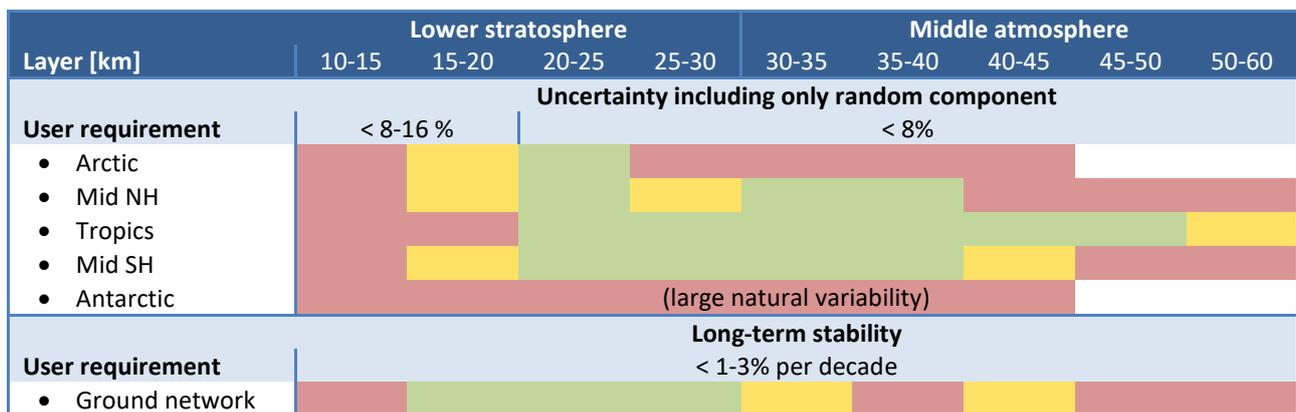


6.4.6.6 Compliance with user requirements

The SCIAMACHY record is nearly compliant with most sampling and resolution requirements. The data quality does not meet the requirements in the lowermost part of the ozone profiles, but improves at higher altitudes. The target on random uncertainty is reached around 15-20 km. Stability is compliant with requirements between 15-30 km and partially in parts of the upper stratosphere. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over large part of the polar atmosphere, even though this may not reflect the actual data quality.

Table 6.11 - Compliance of SCIAMACHY UBr v3.5 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	300–400 km [RD87]
Vertical resolution	< 1–3 km	4–5 km
Observation frequency	< 3 days	6 days
Time period	(1980-2010) – (2003-2010)	08/2002 – 04/2012
Total uncertainty in height attribution	< ± 500 m	± 200 m [RD91; RD31]
Dependences	–	latitude, altitude





6.4.7 Odin OSIRIS v5.10

Identification data record	
Observation principle	UV-visible limb scattering, along track scanning
Platform	Odin, polar orbit, late afternoon sun-synchronous precession
Responsible institute	University of Saskatchewan
Contact person	Chris Roth (chris.roth@usask.ca)
Coverage	
• Time	11/2001 – 01/2020
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 10 – 59 km (50 levels); PRS : 400 – 0.2 hPa (31 levels)
L1 processor and version	v5
L2 processor and version	SaskMART v5.10
Validated L2 file version	ALT : fv0002
Retrieval representation	O3 number density versus geometric altitude

6.4.7.1 Co-locations / Validation sample

Figure 6.23 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with the OSIRIS v5.10 dataset. The co-locations cover most latitude zones and are well distributed over the time series, except the winter hemisphere where OSIRIS cannot observe the dark limb. The Odin platform has also been used for astronomy purposes half of the time at the start of the mission, resulting in less OSIRIS observations until June 2007. In recent years, it operates in a reduced duty cycle as yielding less comparisons since about 2012.

6.4.7.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between OSIRIS v5.10 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.24 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.25 show the same information calculated in 30° latitude zones.

The bias profile has a clear vertical structure. A 5-10% negative bias is seen in the lowermost stratosphere. A sudden 5-10% increase of OSIRIS ozone appears between 20-25 km which leads to positive biases (0-3%) at nearly all latitudes. The bump, also noted in earlier versions of OSIRIS data, is supposedly related to a bias in the aerosol retrieval that serve as input to the ozone retrieval [RD23, RD24]. Between 30-50 km bias is generally positive (except in Arctic) and remains mostly less than 5%.

The general structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Between 20-50 km, the spread is about 5-8% at the equator, and increases gradually to ~10% at the poles. Below 20 km, the comparison spread increases rapidly, reaching 40% around 16 km in the tropics and around 11 km at higher latitudes. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ provided in the OSIRIS record (not shown here). The latter is about 2-3% above 20 km and rapidly increases to ~15% around the tropopause.

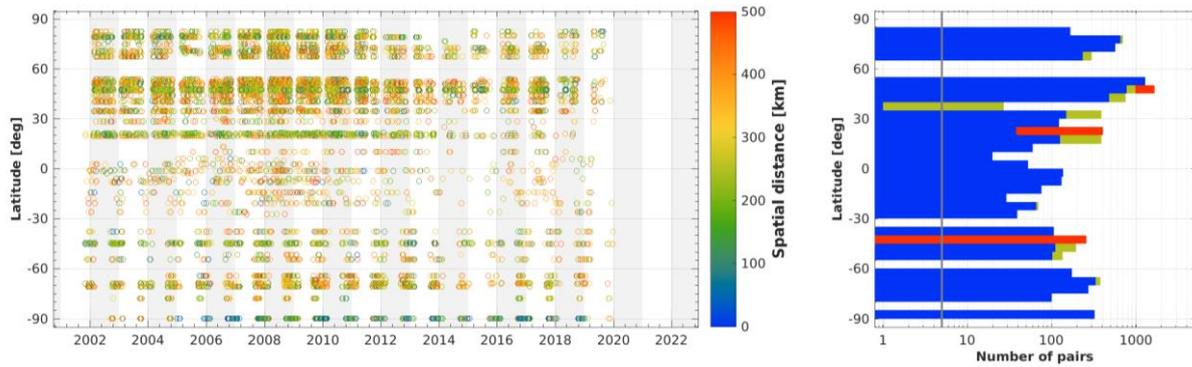


Figure 6.23 - (Left) Latitude–time distribution of co-locations between OSIRIS v5.10 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

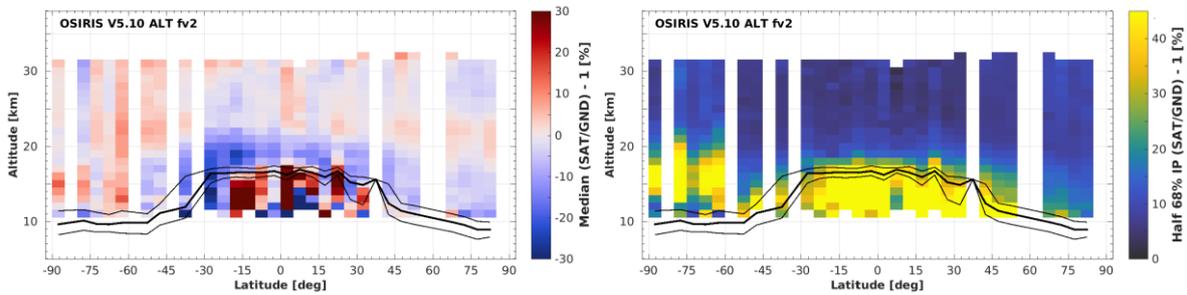


Figure 6.24 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between OSIRIS v5.10 ozone profile data and the global ozonesonde network, calculated over the entire OSIRIS time period and in 5° bins. Black lines indicate the median (thick) and 1 σ spread (thin) of the tropopause altitude.

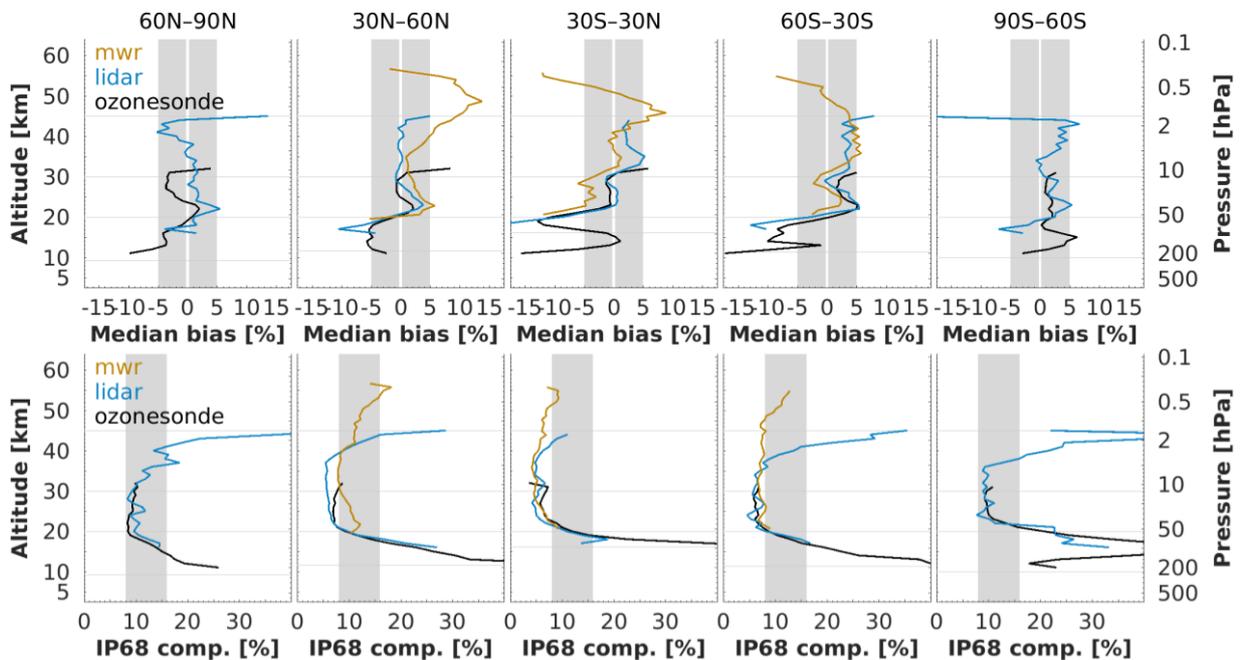


Figure 6.25 - (Top) Median bias between OSIRIS v5.10 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

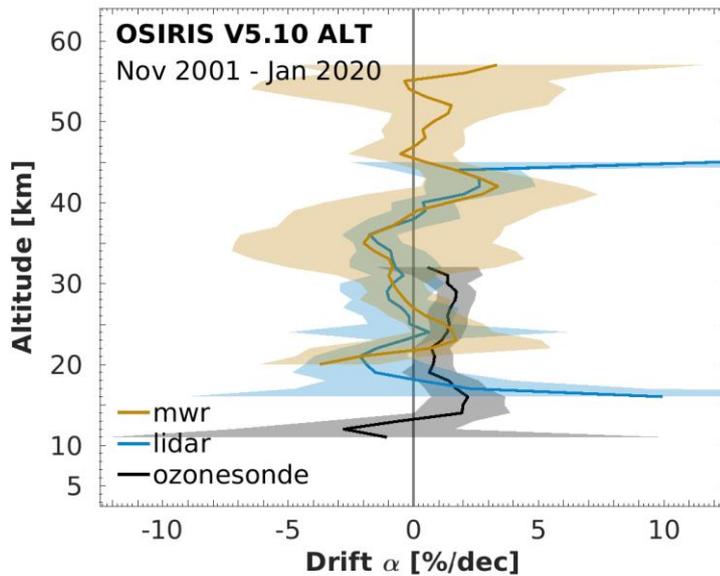


Figure 6.26 - Global average drift (in percent / decade) of OSIRIS v5.10 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2001-2020 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.7.3 Long-term stability

Figure 6.26 shows the vertical structure of the decadal drift of OSIRIS v5.10 with respect to co-located ground-based data [RD28]. Two peaks are clearly visible in the drift profile in the upper stratosphere. A maximum of +2.8% per decade is noted around 42 km, both the lidar and the MWR results are significant. Several km below, around 35 km, comparisons to lidar and MWR indicate a negative drift of OSIRIS by -1.6% per decade, although only the lidar results is significant. Lidar and MWR estimates are generally in very good agreement across the common vertical range. Sonde results diverge somewhat, being 1-2% per decade more positive. This leads, in the middle and lower stratosphere, to 1-2% per decade positive drift of OSIRIS w.r.t. sonde and a negative drift of 1% per decade w.r.t. lidar/MWR. Considering the differing drift estimates between reference instruments, we conclude that the OSIRIS record is stable in this part of the stratosphere.

6.4.7.4 Dependence of data quality on other parameters

Further parameters which possibly impact the OSIRIS v5.10 data quality were studied, including optics temperature, solar zenith angle and surface albedo. There were no clear signs that the bias or comparison spread correlate with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of OSIRIS ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the OSIRIS HARMOZ data files and in the correlative data files are consistent.



6.4.7.5 Summary table of validation results

Table 6.12 - Summary of comparison results between OSIRIS v5.10 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-10, +5]	[-4, +4]	[-5, +4]	–
• Middle latitudes	[-8, 0]	[0, +4]	[-1, +5]	[-5, +12]
• Tropics	[-12, 0]	[-10, 0]	[-3, +5]	[-10, +8]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8/20 – 25/>40	8–12/20	10–30	–
• Middle latitudes	8–40	6–8	6–10	10–15
• Tropics	13 – >45	5–10	5–6	6–9
Range long-term stability (drift; %/decade)				
• Ground network	[-3, +2]	[-1, +2]	[-2, +3]	[0, +3]

6.4.7.6 Compliance with user requirements

The OSIRIS v5.10 data record is nearly compliant with most of sampling and resolution requirements. The long-term stability of the OSIRIS record complies with user requirements over the entire profile. A significant +3% per decade drift was identified around 42 km, but this is within the requirements. Random uncertainty generally meets the requirement between 20 km and the stratosphere, except in the polar regions. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled over the entire vertical range, even though this may not reflect the actual data quality.

Table 6.13 - Compliance of OSIRIS v5.10 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	300–400 km [RD87]
Vertical resolution	< 1–3 km	2–4 km
Observation frequency	< 3 days	less frequent before June 2007 and since 2012
Time period	(1980-2010) – (2003-2010)	11/2001 – 01/2020
Total uncertainty in height attribution	< ± 500 m	< ± 400 m [RD66]
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic	(large natural variability)				(large natural variability)				
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic									
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.8 SCISAT-1 ACE-FTS v3.5/v3.6

The only difference between v3.5 and v3.6 data is the operating system on which the retrieval algorithm runs, resulting numerical differences are small. Below, we refer to this data set as v3.6. Assessment of an extended and updated version (v4.1) of the ACE-FTS data set is reported in the following Section 6.4.9.

Identification data record	
Observation principle	Infrared solar occultation
Platform	SciSat-1, 73.9° inclination orbit
Responsible institute	University of Toronto
Contact person	Patrick Sheese (psheese@atmosph.physics.utoronto.ca)
Coverage	
• Time	02/2004 – 06/2017
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 94 km (89 levels); PRS : 450 – 2×10^{-4} hPa (53 levels)
L1 processor and version	–
L2 processor and version	v3.5, v3.6
Validated L2 file version	ALT : fv0001*; PRS : fv0001
Retrieval representation	O3 volume mixing ratio versus geometric altitude
Related studies	[RD76]

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.8.1 Co-locations / Validation sample

Figure 6.27 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with the ACE-FTS v3.6 dataset. The sampling is quite well distributed over the time series. Due to the 75° inclination orbit and the solar occultation mode it covers mainly the polar regions and the middle latitudes. Although the comparison statistics at tropical sites is limited it is possible to obtain some information by considering the entire tropical belt. Slightly fewer co-locations are also seen in the last few years of the mission, due to the unavailability of publicly released correlative data.

6.4.8.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between ACE-FTS v3.6 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.28 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.29 shows the same information calculated in 30° latitude zones.

ACE-FTS generally overestimates ozone by 3-5% or less from the upper stratosphere down to 10-12 km, except in the Arctic lower stratosphere where the bias may be negative. Between 35-40 km a sudden change of ~5% in bias is noted in the lidar comparisons although the sign is not coherent across all latitudes and a similar feature appears in the MWR results slight lower in the atmosphere. Above the stratopause a large, positive bias relative to MWR starts to develop.

The comparison spread $s_{\Delta x}$ has similar behaviour as for most HARMOZ data sets. In the middle and upper stratosphere the spread is 3-5% at mid latitudes and increases to 6-8% towards the poles. In the Antarctic the variability is somewhat larger due to the occurrence of the ozone hole. Below 20 km the comparison spread increases rapidly, reaching 40% at the tropopause. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ provided in the ACE-FTS record (not shown here). The latter is of the order of 1-2% above 20 km and increase up to 10-15% in the UTLS and upper troposphere.

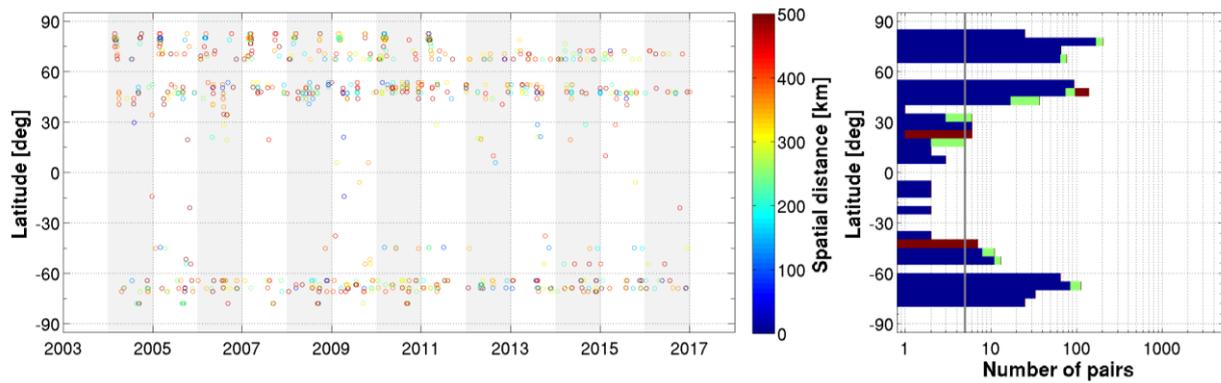


Figure 6.27 - (Left) Latitude–time distribution of co-locations between ACE-FTS v3.6 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

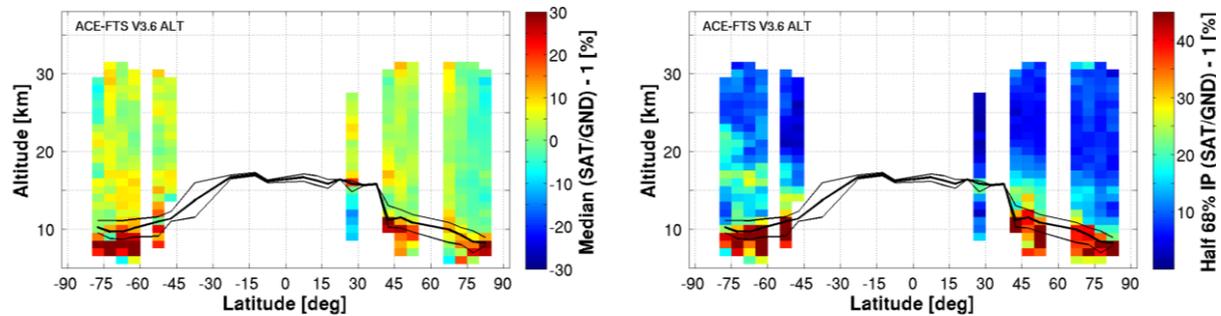


Figure 6.28 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between ACE-FTS v3.6 ozone profile data and the global ozonesonde network, calculated over the entire ACE-FTS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

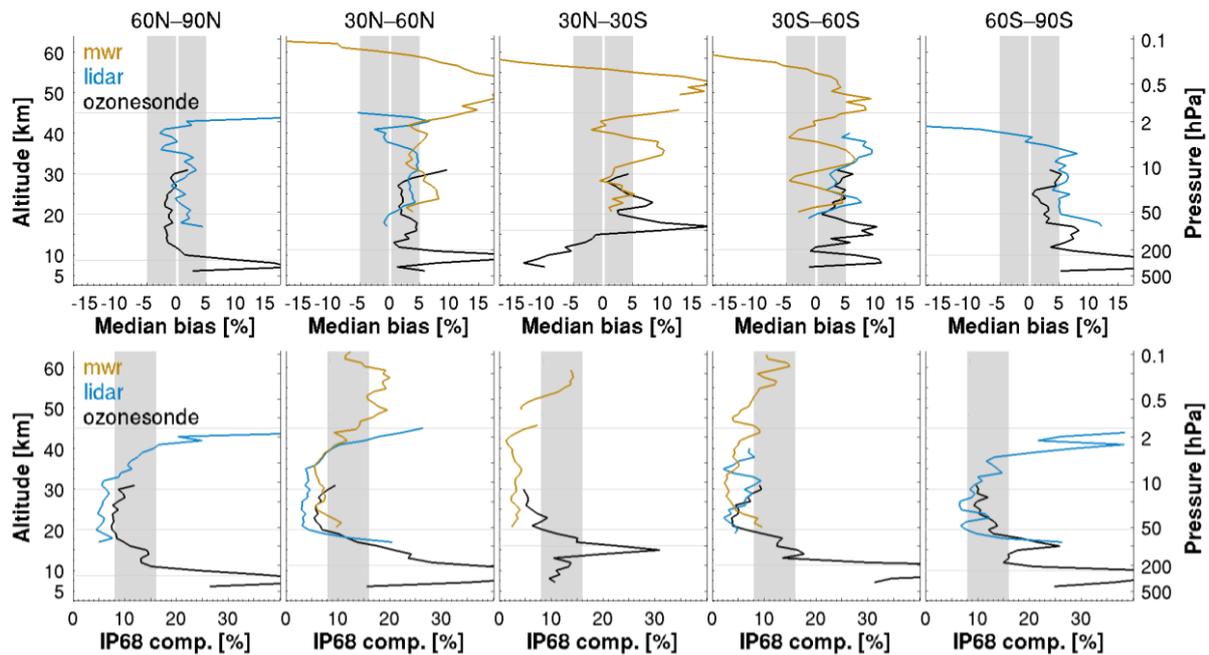


Figure 6.29 - (Top) Median bias between ACE-FTS v3.6 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

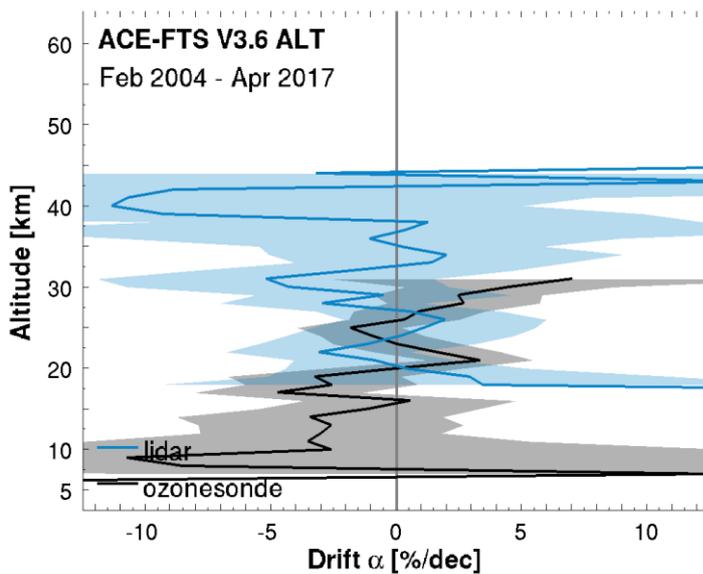


Figure 6.30 - Global average drift (in percent / decade) of ACE-FTS v3.6 ozone profile data with respect to co-located ozonesonde (black) and lidar (blue) network data, calculated over the 2004-2017 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.8.3 Long-term stability

Figure 6.30 shows the vertical structure of the decadal drift of ACE-FTS v3.6 data relative to the ozonesonde and lidar networks. The number of co-locations with lidar measurements is limited, so the analysis is not as sensitive to drift as compared to other HARMOZ data sets. ACE-FTS drift estimates are negative over most of the stratosphere, though never significant. Between 10-35 km the average drift value is $-(1-3)\%$ per decade. In the uppermost stratosphere larger values are found but larger drift uncertainty as well. Nonetheless, we would advise caution when using ACE-FTS data in this part of the stratosphere. The poorer sampling



properties of the solar occultation technique intrinsically leads to fewer co-locations and hence a more limited assessment of stability. It is therefore difficult to detect drift of ACE-FTS by less than 4% per decade.

6.4.8.4 Dependence of data quality on other parameters

Further parameters which possibly impact the ACE-FTS v3.6 data quality were studied, such as the beta angle of the occultation. There were no clear signs that the bias or comparison spread correlate with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of ACE-FTS ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the ACE-FTS HARMOZ data files and in the correlative data files are consistent.

6.4.8.5 Summary table of validation results

Table 6.14 - Summary of comparison results between ACE-FTS v3.6 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-2, +8]	[-2, +5]	[-10, +5]	–
• Middle latitudes	[0, +9]	[+2, +5]	[-2, +7]	[-8, +7]
• Tropics	[-11, +15]	[+1, +7]	[-1, +10]	[-10, +16]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8–28	6–7/10	6–35	–
• Middle latitudes	7–40	4–7	4–12	12–20
• Tropics	8–32	5–8	3–5	4–14
Range long-term stability (drift; %/decade)				
• Ground network	[-4, 0]	[-3, +2]	[-11, +2]	–



6.4.8.6 Compliance with user requirements

The ACE-FTS v3.5/v3.6 ozone profile data record is compliant with most sampling and resolution requirements. The data quality was assessed against ground-based measurements in the polar regions and at northern middle latitudes. It generally meets the user requirements between 20 km and the stratopause. Stability is compliant with requirements in the middle stratosphere and perhaps not in the upper stratosphere though lidar results are not significant. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled in the Antarctic, even though this may not reflect the actual data quality.

Table 6.15 - Compliance of ACE-FTS v3.6 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain
Vertical resolution	< 1–3 km	3 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	02/2004 – 06/2017
Total uncertainty in height attribution	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
User requirement	Uncertainty including only random component								
	< 8-16 %				< 8%				
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 									
User requirement	Long-term stability								
	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.9 SCISAT-1 ACE-FTS v4.1

Assessment of an earlier version (v3.5/v3.6) of the ACE-FTS data set has been reported in the preceding Section 6.4.8.

Identification data record	
Observation principle	Infrared solar occultation
Platform	SciSat-1, 73.9° inclination orbit
Responsible institute	University of Toronto
Contact person	Patrick Sheese (psheese@atmosph.physics.utoronto.ca)
Coverage	
• Time	02/2004 – 12/2020
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 94 km (89 levels)
L1 processor and version	–
L2 processor and version	v4.1
Validated L2 file version	ALT : fv0001
Retrieval representation	O3 volume mixing ratio versus geometric altitude
Related studies	[RD77]

6.4.9.1 Co-locations / Validation sample

Figure 6.31 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with the ACE-FTS v4.1 dataset. The sampling is quite well distributed over the time series. Due to the 75° inclination orbit and the solar occultation mode mainly the polar regions and the middle latitudes are covered. Although the comparison statistics at tropical sites is limited it is possible to obtain some information by considering the entire tropical belt.

6.4.9.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between ACE-FTS v4.1 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.32 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.33 shows the same information calculated in 30° latitude zones.

ACE-FTS v4.1 biases are generally positive and exhibit a pronounced vertical structure. Local minima are noted around 17-20 km (0%) and around 35-42 km (+5-7%). Local maxima are seen close to the tropopause (+7-10%), around 30 km (+8-12%) and around 50-55 km (+13-20%). Biases may become negative in the polar upper stratosphere, although constraints are challenging in this region due to the limited number of lidar sites and their operation is limited to polar night.

The comparison spread $s_{\Delta x}$ has similar behaviour as for most HARMOZ data sets. In the middle and upper stratosphere the spread is 3-5% at mid latitudes and increases to 6-8% towards the poles. In the Antarctic the variability is somewhat larger due to the occurrence of the ozone hole. Below 20 km, the comparison spread increases rapidly, reaching 40% at the tropopause. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ provided in the ACE-FTS record (not shown here). The latter is of the order of 1-2% above 20 km and increase up to 10-15% in the UTLS and upper troposphere.

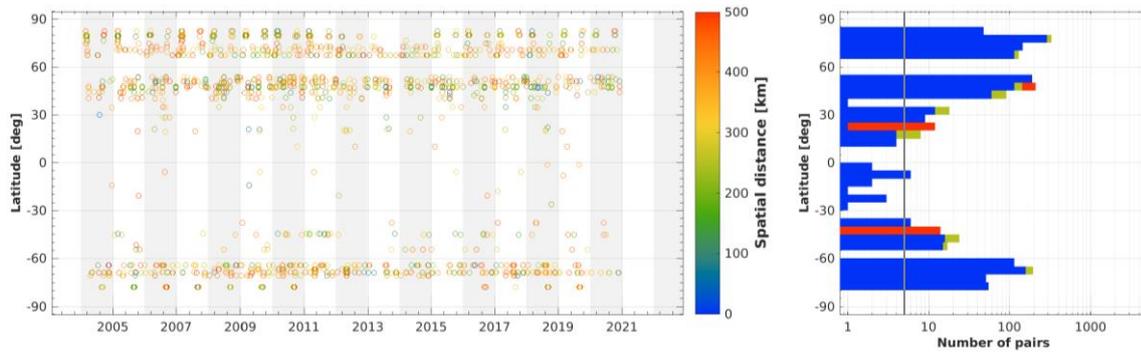


Figure 6.31 - (Left) Latitude–time distribution of co-locations between ACE-FTS v4.1 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

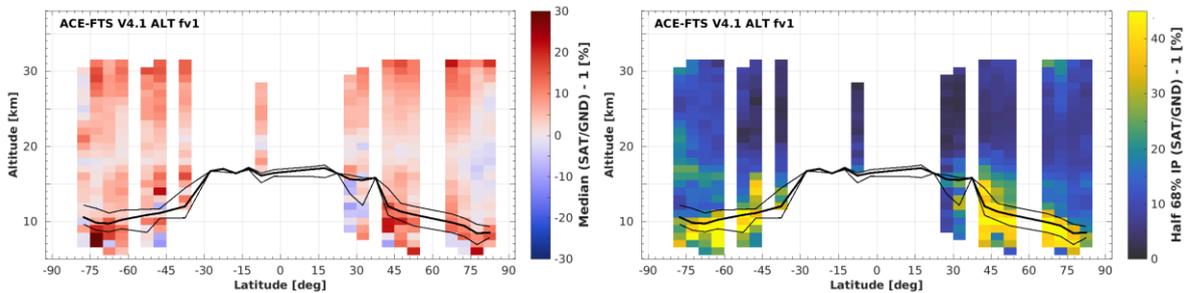


Figure 6.32 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between ACE-FTS v4.1 ozone profile data and the global ozonesonde network, calculated over the entire ACE-FTS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

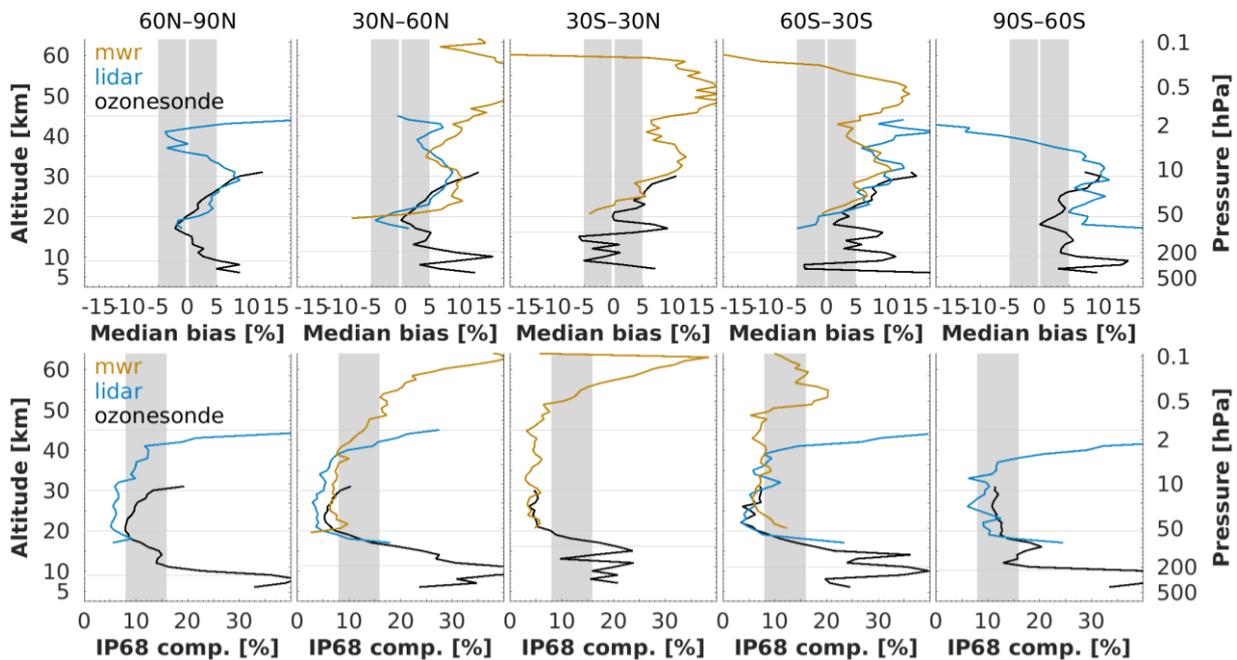


Figure 6.33 - (Top) Median bias between ACE-FTS v4.1 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

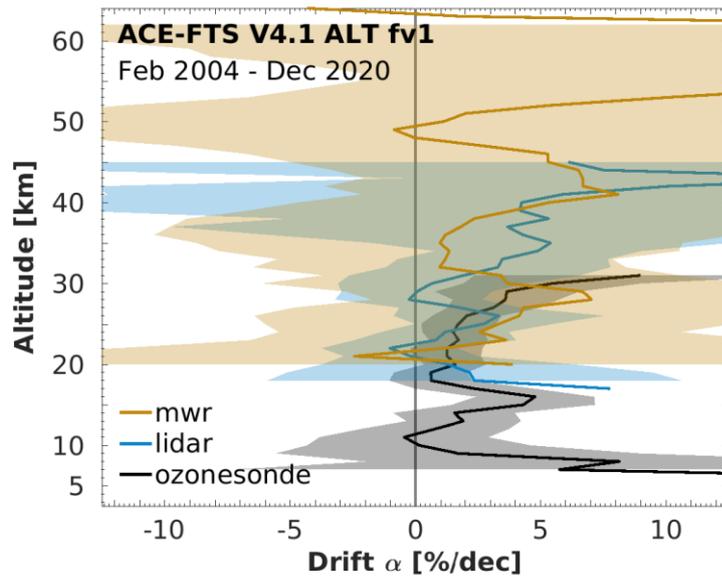


Figure 6.34 - Global average drift (in percent / decade) of ACE-FTS v4.1 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2004-2020 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.9.3 Long-term stability

Figure 6.34 shows the vertical structure of the decadal drift of ACE-FTS v4.1 data relative to the ozonesonde, lidar and MWR networks. The number of co-locations with lidar measurements is limited, so the analysis is less sensitive to drift as compared to other HARMOZ data sets. ACE-FTS drift estimates are positive over most of the stratosphere. Only the ozonesonde results are statistically significant. Between 10-35 km the average drift value is +(1-3)% per decade. In the uppermost stratosphere larger values (4-6% per decade) are found but the uncertainty is larger as well. Since two independent ground-based records (lidar & MWR) lead to similar findings, we advise caution when using ACE-FTS data in this part of the stratosphere. The poorer sampling properties of the solar occultation technique intrinsically leads to fewer co-locations and hence a more limited assessment of stability. It is therefore difficult to detect drift of ACE-FTS by less than 2-3% per decade.

6.4.9.4 Dependence of data quality on other parameters

Further parameters which possibly impact the ACE-FTS v4.1 data quality were studied, such as the beta angle of the occultation. There were no clear signs that the bias or comparison spread correlate with any of these parameters. Any apparent dependence turned out to originate mainly from the correlation of the parameter with altitude and latitude.

The bias, the comparison spread and the long-term stability of ACE-FTS ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the ACE-FTS HARMOZ data files and in the correlative data files are consistent.



6.4.9.5 Summary table of validation results

Table 6.16 - Summary of comparison results between ACE-FTS v4.1 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-2, +5]	[0, +10]	[-10, +10]	–
• Middle latitudes	[0, +7]	[0, +11]	[+4, +12]	[-5, +20]
• Tropics	[-3, +10]	[0, +9]	[+7, +11]	[-5, +16]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8–28	6–7/10	6–35	–
• Middle latitudes	6–40	4–7	5–12	8–25
• Tropics	8–20	4–6	3–6	5–28
Range long-term stability (drift; %/decade)				
• Ground network	[0, +4]	[+1, +3]	[+2, +6]	–

6.4.9.6 Compliance with user requirements

The ACE-FTS v4.1 ozone profile data record is compliant with most sampling and resolution requirements. The data quality was assessed against ground-based measurements at high and middle latitudes. It generally meets the user requirements between 20 km and the stratopause. Stability is compliant with requirements in the lower and middle stratosphere and likely not compliant in the upper stratosphere. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled in the Antarctic, even though this may not reflect the actual data quality.

Table 6.17 - Compliance of ACE-FTS v4.1 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain
Vertical resolution	< 1–3 km	3 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	02/2004 – 12/2020
Total uncertainty in height attribution	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic									
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic									
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.10 Suomi-NPP OMPS-LP USask-2D v1.0.2

Assessment of an extended and updated version (v1.1.0) of the OMPS-LP data set is reported in the following Section 6.4.11.

Identification data record	
Observation principle	UV-visible limb scattering
Platform	Suomi NPP, afternoon polar orbit, sun-synchronous precession
Responsible institute	University of Saskatchewan
Contact person	Daniel Zawada (daniel.zawada@usask.ca)
Coverage	
• Time	01/2012 – 02/2017
• Latitude	82°S – 82°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 58 km (53 levels), PRS : 400 – 0.2 hPa (31 levels)
L1 processor and version	v2.0-v2.4 – center slit
L2 processor and version	USask-2D v1.0.2 – center slit
Validated L2 file version	ALT: fv0004*; PRS : fv0002
Retrieval representation	O3 number density versus geometric altitude

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.10.1 Co-locations / Validation sample

Figure 6.35 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with OMPS-LP USask-2D v1.0.2 data. The sampling covers most latitude zones and is homogeneous in time, except for a short interruption at the end of 2013 and in the polar winter atmosphere.

6.4.10.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between OMPS-LP v1.0.2 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC microwave radiometers are shown in the figures below. Figure 6.36 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.37 shows the same information calculated in 30° latitude zones.

Sonde and lidar comparisons show a coherent picture of OMPS bias structure in the middle and lower stratosphere. Bias is generally smaller than 5% but a clear 10 km-wide peak is centred around 18-20 km. Around the tropopause there is a systematic underestimation of 10-15% at all latitudes. In the upper stratosphere a positive bias of 3-8% is noted relative to lidar which agrees broadly with the MWR results. The bias results in the mesosphere are very scattered and no clear picture emerges.

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above 20-25 km the half-IP-68 spread is about 4% at the equator and increases gradually up to 6-10% at the poles. Below ~20 km the spread increases rapidly, reaching 25-30% at the tropopause. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{ex-ante}$ (not shown here) provided in the OMPS-LP data files. The latter is of the order of 2-3% above ~20 km and rapidly increase up to 10-15% at the tropopause.

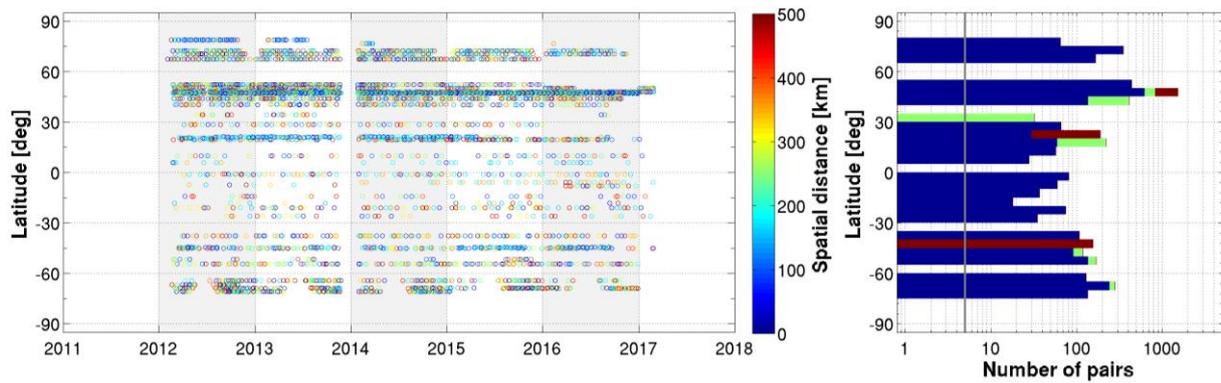


Figure 6.35 - (Left) Latitude–time distribution of co-locations between OMPS-LP USask-2D v1.0.2 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

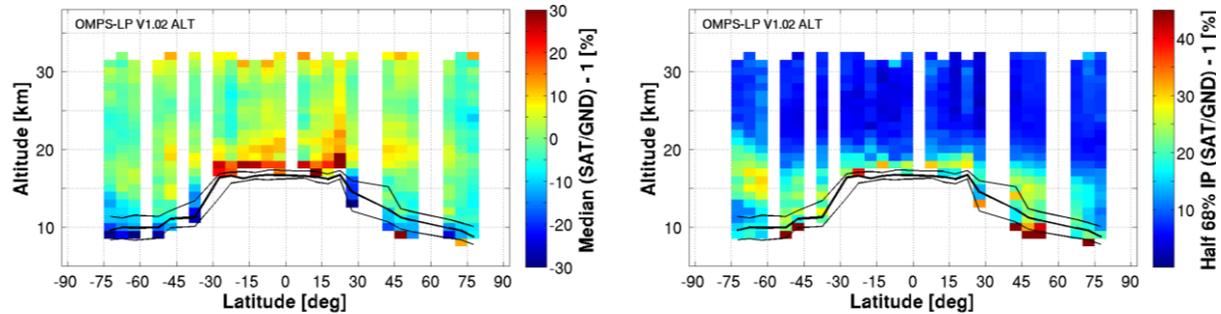


Figure 6.36 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between OMPS-LP USask-2D v1.0.2 ozone profile data and the global ozonesonde network, calculated over the entire OMPS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

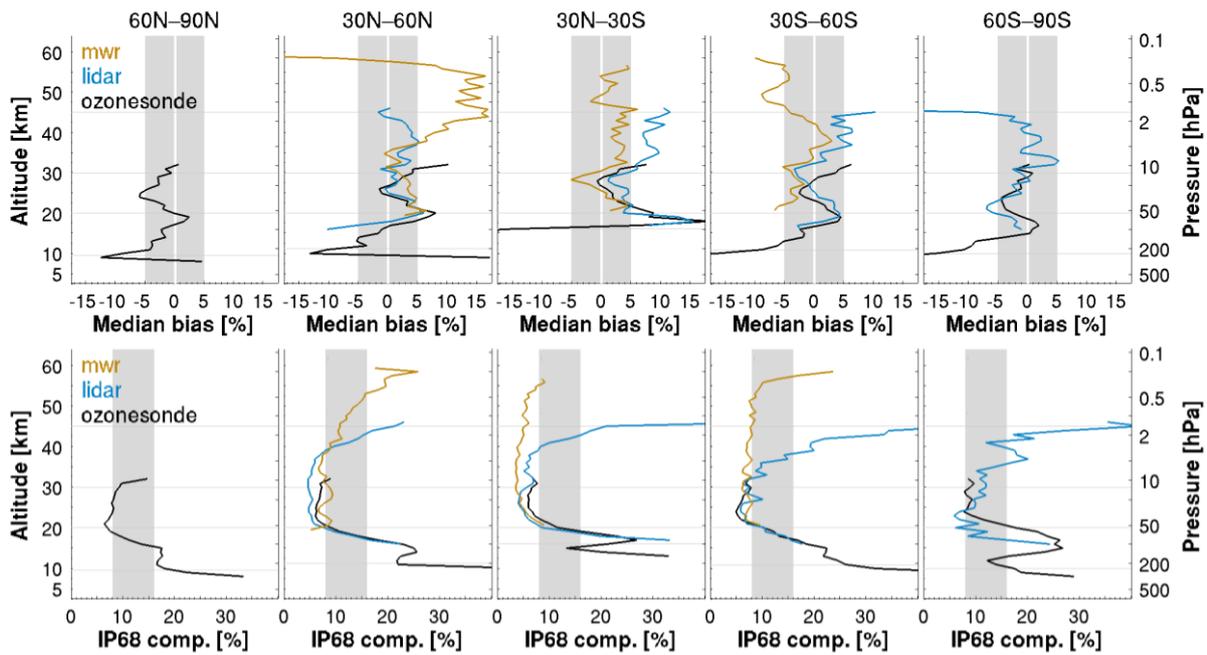


Figure 6.37 - (Top) Median bias between OMPS-LP USask-2D v1.0.2 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

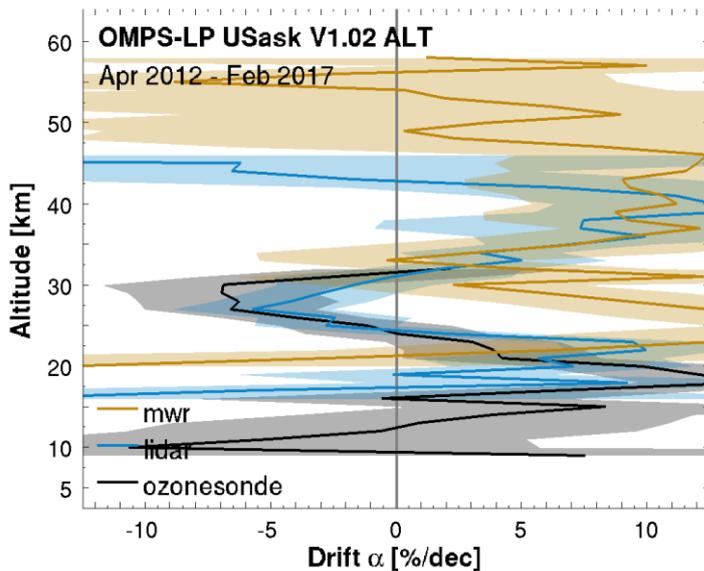


Figure 6.38 - Global average drift (in percent / decade) of OMPS-LP USask-2D v1.0.2 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2012-2017 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.10.3 Long-term stability

Figure 6.38 shows the drift of OMPS-LP USask-2D v1.0.2 ozone profile data with respect to co-located ozonesonde, lidar and MWR network data. Though the record is only five years long, the ample co-location statistics allows to estimate drift with comparable precision (1.5% per decade, 1 σ) as that for many other limb/occultation sounders in this report. A fairly coherent picture emerges for all ground-based techniques, except in regions where these are less sensitive (MWR below 30 km, lidar close to the stratopause).



The OMPS-LP drift profile oscillates between negative values (-6% per decade) at ~27 km and positive values at ~18 km (+7% per decade) and ~40 km (+9% per decade). Although drift uncertainty may be underestimated due to the shorter time period the good agreement between independent results from different ground-based data records builds confidence in attributing observed drift to OMPS-LP. As a consequence, we advise great care in using these data for studies of long-term changes in ozone.

We also note that this version of the Level-2 data is retrieved from version 2.0-2.4 calibrated Level-1 data which has large uncertainties in the altitude registration (Moy et al., 2017; RD68). An improved pointing correction scheme was recently released (Level-1 v2.5) and this impacted the structure of drift in the ozone profile records substantially (Kramarova et al., 2017, RD54).

6.4.10.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of OMPLS-LP USask-2D v1.0.2 are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the OMPS-LP HARMOZ data files and in the correlative data files are consistent.

6.4.10.5 Summary table of validation results

Table 6.18 - Summary of comparison results between OMPS-LP USask v1.0.2 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-13, +2]	[-6, 0]	[-5, +5]	–
• Middle latitudes	[-8, +8]	[-2, +5]	[-4, +5]	[-10, +15]
• Tropics	[+8, +18]	[0, +8]	[+3, +9]	[-2, +5]
Range comparison spread (half 68% interpercentile; %)				
• Polar	7–17 / 12–27	6–12	12–25	–
• Middle latitudes	8–25	5–8	7–10	8–23
• Tropics	12–27	4–7	3–5	5–9
Range long-term stability (drift; %/decade)				
• Ground network	[-9, +11]	[-7, +10]	[-4, +10]	[-6, +10]



6.4.10.6 Compliance with user requirements

From the results reported above it can be concluded that the OMPS-LP USask-2D v1.0.2 ozone profile data record is nearly compliant with most sampling and resolution requirements. Random uncertainty meets the requirements nearly everywhere between 20-60 km. However, the long-term drift exceeds the threshold over most of the profile, likely as a result of pointing inaccuracies. Also, assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled in the Antarctic, but it is understood that this may not reflect the actual data quality.

Table 6.19 - Compliance of OMPS-LP USask v1.0.2 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	250-400 km
Vertical resolution	< 1–3 km	1–2 km
Observation frequency	< 3 days	3-4 days [RD54]
Time period	(1980-2010) – (2003-2010)	01/2012 – 02/2017
Total uncertainty in height registration	< ± 500 m	~400 m [RD68]
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic									
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic					(large natural variability)				
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.11 Suomi-NPP OMPS-LP USask-2D v1.1.0

Assessment of an earlier version (v1.0.2) of the OMPS-LP data set has been reported in the preceding Section 6.4.10.

Identification data record	
Observation principle	UV-visible limb scattering
Platform	Suomi NPP, afternoon polar orbit, sun-synchronous precession
Responsible institute	University of Saskatchewan
Contact person	Daniel Zawada (daniel.zawada@usask.ca)
Coverage	
• Time	01/2012 – 01/2020
• Latitude	82°S – 82°N
• Longitude	180°W – 180°E
• Vertical	ALT : 6 – 58 km (53 levels)
L1 processor and version	v2.5 – center slit
L2 processor and version	USask-2D v1.1.0 – center slit
Validated L2 file version	ALT: fv0001
Retrieval representation	O3 number density versus geometric altitude

6.4.11.1 Co-locations / Validation sample

Figure 6.39 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with OMPS-LP USask-2D v1.1.0 data. The sampling covers nearly all latitude belts and it is homogeneous in time, except for a short interruption at the end of 2013 and in the polar winter atmosphere.

6.4.11.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between OMPS-LP v1.1.0 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC microwave radiometers are shown in the figures below. Figure 6.40 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.41 shows the same information calculated in 30° latitude zones.

Sonde and lidar comparisons show a coherent picture of OMPS bias structure in the middle and lower stratosphere. Bias is generally smaller than 5% but a clear 10 km-wide peak is centred around 18-20 km. Around the tropopause there is a systematic underestimation of 10-15% at all latitudes. In the upper stratosphere a positive bias of 5-10% is noted relative to lidar which agrees broadly with the MWR results. The bias results in the mesosphere are very scattered and no clear picture emerges.

We find a change in the vertical gradient of the bias profiles w.r.t. ground-based data from OMPS-LP version 1.0.2 (Figure 6.37, top) to v1.1.0 (Figure 6.41, top). Biases in v1.1.0 data are generally more negative than in v1.0.2 data below ~20 km and, above this level, the biases become more positive in the recent data version. The change in gradient and the location of unchanged bias (~20 km, around the peak in the ozone number density profile) suggests an upward shift of the altitude registration of the vertical profile in the v1.1.0 data version (see also Section 6.4.11.3).

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above 20-25 km the half-IP-68 spread is about 4% at the equator and increases gradually up to 6-10% at the poles. Below ~20 km the spread increases rapidly, reaching 25-30% at the tropopause. The spread seen in the comparisons is a few percent larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ (not shown here) provided in the OMPS-LP



data files. The latter is of the order of 2-3% above ~20 km and rapidly increase up to 10-15% at the tropopause. These results are virtually identical to those of v1.0.2.

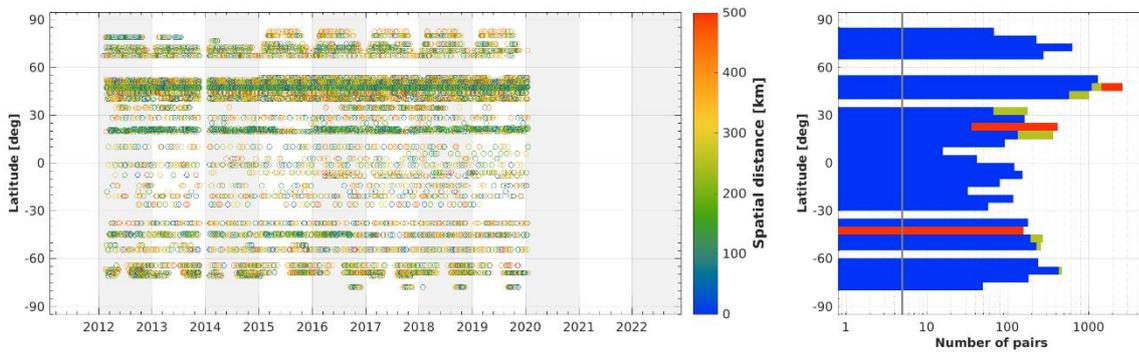


Figure 6.39 - (Left) Latitude–time distribution of co-locations between OMPS-LP USask-2D v1.1.0 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

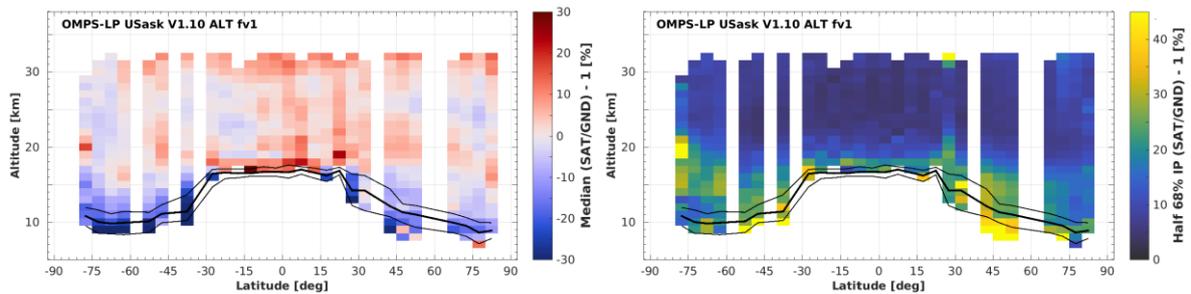


Figure 6.40 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between OMPS-LP USask-2D v1.1.0 ozone profile data and the global ozonesonde network, calculated over the entire OMPS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

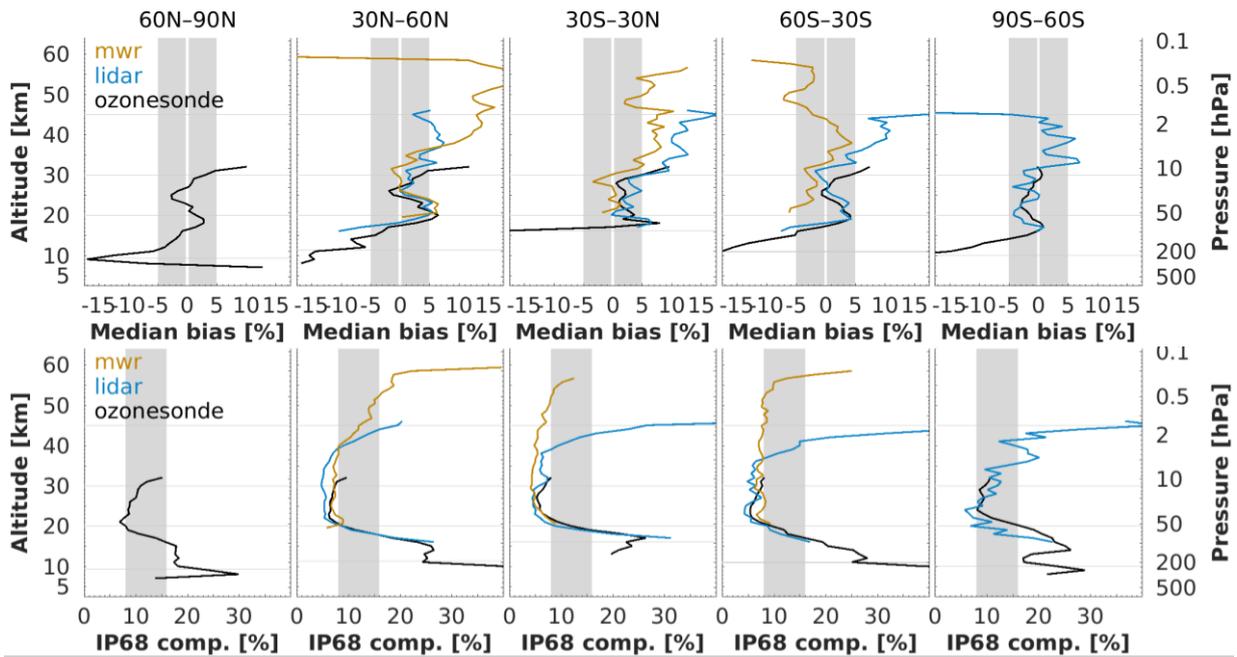


Figure 6.41 - (Top) Median bias between OMPS-LP USask-2D v1.1.0 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

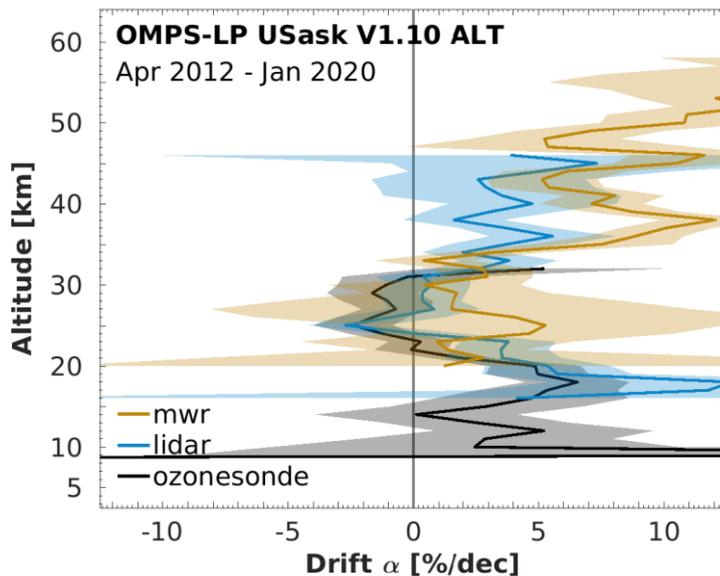


Figure 6.42 - Global average drift (in percent / decade) of OMPS-LP USask-2D v1.1.0 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2012-2020 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.11.3 Long-term stability

Figure 6.42 shows the drift of OMPS-LP USask-2D v1.1.0 ozone profile data with respect to co-located ozonesonde, lidar and MWR network data. The ample co-location statistics allows to estimate drift with a precision of about 1-1.5% per decade (1 σ). A coherent picture emerges for all ground-based techniques.

The OMPS-LP drift profile oscillates between insignificant negative values (-1% per decade) at ~27 km and statistically significant positive values in the lower stratosphere (+5% per decade at ~18 km) and the upper stratosphere (+4% per decade at ~40 km). MWR comparisons suggest that the positive drift extends into the



lower mesosphere. The good agreement between independent results from different ground-based data records builds confidence in attributing the observed drift and its vertical structure to OMPS-LP. As a consequence, we advise great care in using these data for studies of long-term changes in ozone.

The Level-2 v1.1.0 profiles were retrieved from Level-1 v2.5 data with improved altitude registration with respect to the previous OMPS-LP data version (Kramarova et al., 2017, RD54). This change improves the stability and its vertical dependence of the ozone profile record as can be seen when comparing Figure 6.38 and Figure 6.42. Even though v1.1.0 is more stable than its predecessor, the OMPS-LP data record continues to exhibit a clear drift in the lowermost stratosphere, in the upper stratosphere and possibly in the lower mesosphere as well. Hence, in large regions of the atmosphere this data set fails to comply with user requirements (1-3% per decade) and further work will be needed to address this issue.

6.4.11.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of OMPLS-LP USask-2D v1.1.0 are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the OMPS-LP HARMOZ data files and in the correlative data files are consistent.

6.4.11.5 Summary table of validation results

Table 6.20 - Summary of comparison results between OMPS-LP USask v1.1.0 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-15, +2]	[-3, +2]	[0, +5]	–
• Middle latitudes	[-12, +5]	[-2, +6]	[-1, +10]	[-6, +17]
• Tropics	[-10, +8]	[0, +3]	[+5, +13]	[-2, +10]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8–18 / 10–25	7–14	12–25	–
• Middle latitudes	8–25	5–8	7–11	8–25
• Tropics	12–25	4–8	4–6	6–12
Range long-term stability (drift; %/decade)				
• Ground network	[+1, +6]	[-2, 0]	[+1, +6]	> +6



6.4.11.6 Compliance with user requirements

From the results reported above it we conclude that the OMPS-LP USask-2D v1.1.0 ozone profile data record is nearly compliant with most sampling and resolution requirements. Random uncertainty meets the requirements almost everywhere between 20-60 km. However, the long-term drift exceeds the threshold over large parts of the profile, likely resulting from remaining of instability in the altitude registration. Also, assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled in the Antarctic, but it is understood that this may not reflect the actual data quality.

Table 6.21 - Compliance of OMPS-LP USask v1.1.0 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	250-400 km
Vertical resolution	< 1–3 km	1–2 km
Observation frequency	< 3 days	3-4 days [RD54]
Time period	(1980–2010) – (2003–2010)	01/2012 – 01/2020
Total uncertainty in height registration	< ± 500 m	~400 m [RD68]
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere					
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60	
Uncertainty including only random component										
User requirement	< 8-16 %				< 8%					
• Arctic										
• Mid NH										
• Tropics										
• Mid SH										
• Antarctic					(large natural variability)					
Long-term stability										
User requirement	< 1-3% per decade									
• Ground network										



6.4.12 ERBS SAGE II v7.0

Identification data record	
Observation principle	UV-visible solar occultation
Platform	ERBS, 56.9° inclination orbit
Responsible institutes	NASA Langley Research Center, University of Bremen – IUP
Contact persons	Robert Damadeo (robert.damadeo@nasa.gov), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	10/1984 – 08/2005
• Latitude	80°S – 80°N
• Longitude	180°W – 180°E
• Vertical	ALT : 0 – 70 km (71 levels); PRS : 1000 – 0.05 hPa (41 levels)
L1 processor and version	–
L2 processor and version	v7.00
Validated L2 file version	ALT : fv0006*; PRS : fv0006
Retrieval representation	O3 number density versus geometric altitude

* All results below are for the ALT data files. There is no notable difference with the PRS data files.

6.4.12.1 Co-locations / Validation sample

Figure 6.43 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with SAGE II v7.0 data. Very few comparisons are available in the tropics, the sample mainly covers middle and high latitudes. It is fairly homogeneous in time, but increases of ground network size are clear in the 1990s.

6.4.12.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between SAGE II v7.0 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.44 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.45 shows the same information calculated in 30° latitude zones.

The SAGE II bias is negative at all latitudes below 20 km. At higher altitudes (20–45 km) the bias is mostly less than 3–4% when compared to all three ground-based techniques. Although some tension arises between sonde and lidar comparison results in polar regions, this can be traced to either poor co-location statistics and hence larger uncertainties on the bias estimates (Arctic), or to a known negative bias in the lidar data between 1991 and 1998 at Dumont d’Urville (Antarctica, [RD41]). In the mesosphere SAGE II reports larger ozone values than MWR instruments, by 5–10%.

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above 20 km the half-IP-68 spread is 4–6% at the equator and increases gradually up to ~10% at the poles. The comparison spread increases rapidly at lower altitudes and reaches 40% around the tropopause. The spread seen in the comparisons is clearly larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ (not shown here) provided in the SAGE II data files. The latter is only 1–2% above 20 km and increases to about 20% at the tropopause.

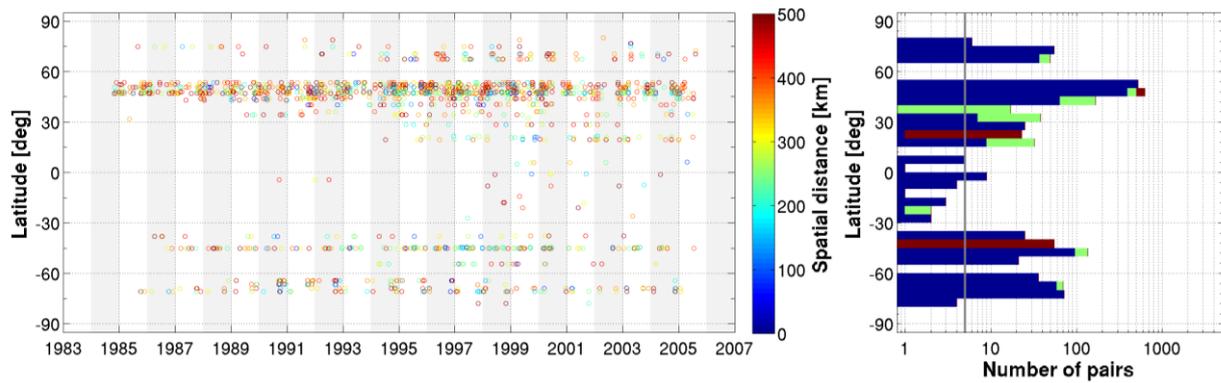


Figure 6.43 - (Left) Latitude–time distribution of co-locations between SAGE II v7 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

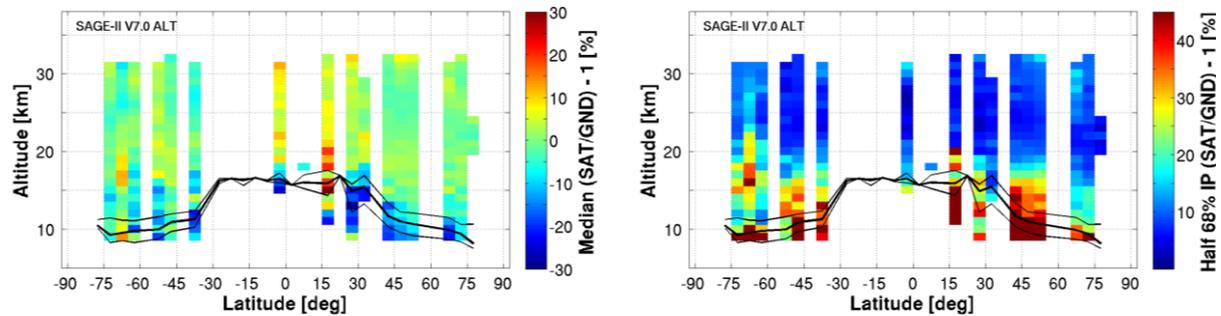


Figure 6.44 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between SAGE II v7 ozone profile data and the global ozonesonde network, calculated over the entire SAGE II time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

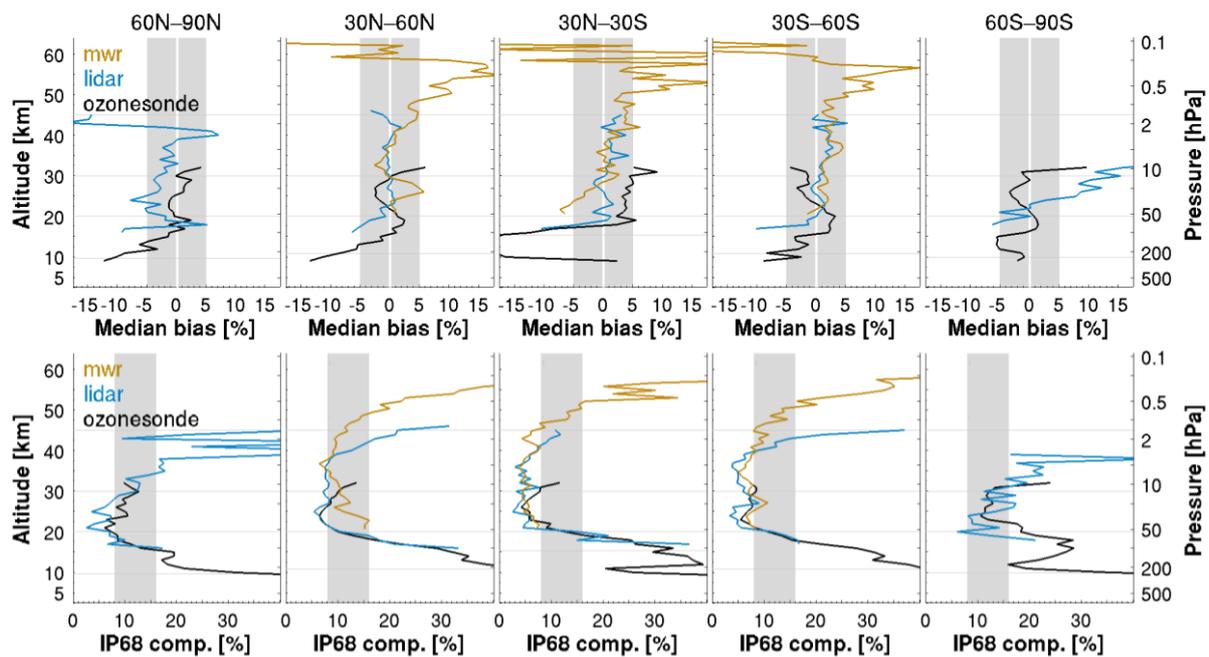


Figure 6.45 - (Top) Median bias between SAGE II v7 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

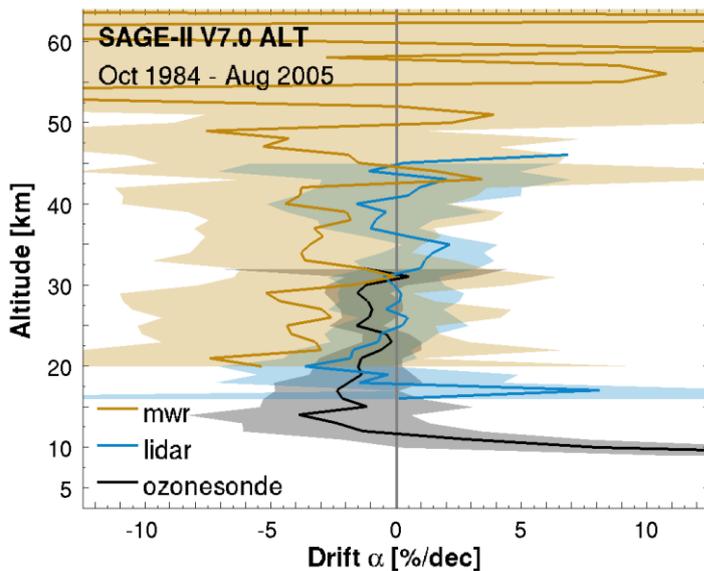


Figure 6.46 - Global average drift (in percent / decade) of SAGE II v7 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 1984-2005 time period (sonde, lidar) or the 1995-2005 period (MWR). The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.12.3 Long-term stability

Figure 6.46 shows the vertical structure of the decadal drift of SAGE II v7.0 data relative to the ozonesonde, lidar and MWR networks. Drift estimates are virtually independent of altitude and they remain between -2% and 1% per decade over the entire stratosphere. Estimates of drift relative to MWR are based on a different period in time than sonde and lidar since regular MWR observations started mostly after 1995. This may explain the quantitative differences in drift, yet the qualitative conclusion remains the same. SAGE II drift



values are small and not significant, hereby showing the excellent stability of the SAGE II data record for studies of long-term changes in ozone.

6.4.12.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of SAGE II ozone are in excellent agreement in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the SAGE II HARMOZ data files and in the correlative data files are consistent.

6.4.12.5 Summary table of validation results

Table 6.22 - Summary of comparison results between SAGE II v7 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-10, +1]	[-7, +2]	[-3, +6]	–
• Middle latitudes	[-8, +2]	[-3, +1]	[-2, +3]	[0, +10]
• Tropics	[-20, +3]	[-1, +4]	[0, +3]	[+2, +10]
Range comparison spread (half 68% interpercentile; %)				
• Polar	8–30	6–12/18	12–28	–
• Middle latitudes	10–40	4–8	5–11	10–35
• Tropics	12–35	4–6	5–7	7–26
Range long-term stability (drift; %/decade)				
• Ground network	[-3, -2]	[-2, 0]	[-3, +2]	[-8, +5]



6.4.12.6 Compliance with user requirements

From the results reported above it can be concluded that the SAGE II v7 ozone profile data record is compliant with most sampling and resolution requirements. Data quality meets the random uncertainty requirements between 20 km and the stratopause, except in polar regions where random mismatch uncertainty due to natural variability obfuscates the analysis results. Decadal stability is fully compliant between 15-45 km and partially at other levels.

Table 6.23 - Compliance of SAGE II v7.0 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain [RD88]
Vertical resolution	< 1–3 km	1 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	10/1984 – 08/2005
Total uncertainty in height registration	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude, sunset/sunrise

Layer [km]	Lower stratosphere				Middle atmosphere					
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60	
Uncertainty including only random component										
User requirement	< 8-16 %				< 8%					
• Arctic										
• Mid NH										
• Tropics										
• Mid SH										
• Antarctic			(large natural variability)							
Long-term stability										
User requirement	< 1-3% per decade									
• Ground network										



6.4.13 UARS HALOE v19

Identification data record	
Observation principle	Infrared solar occultation
Platform	UARS, 57.0° inclination orbit
Responsible institutes	Hampton University, University of Bremen – IUP
Contact persons	James Russell III (james.russell@hamptonu.edu), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	10/1991 – 09/2005
• Latitude	80°S – 80°N
• Longitude	180°W – 180°E
• Vertical	ALT : 1 – 82 km (82 levels); PRS : 500 – 0.02 hPa (40 levels)
L1 processor and version	–
L2 processor and version	v19
Validated L2 file version	ALT : fv0002; PRS : fv0004*
Retrieval representation	O3 volume mixing ratio versus pressure

* All results below are for the PRS data files. There is no notable difference with the ALT data files.

6.4.13.1 Co-locations / Validation sample

Figure 6.47 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with HALOE v19 data. The sampling covers most latitude zones and is reasonably homogeneous in time.

6.4.13.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between HALOE v19 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.48 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.49 shows the same information calculated in 30° latitude.

HALOE underestimates ozone by 10-15% and more below 20-50 hPa (~15-20 km). The bias remains negative but is much smaller, less than 3-4%, in the middle stratosphere. In the upper stratosphere the sign remains less than ~4% and its sign varies with latitude and altitude. Above the stratopause MWR comparisons indicate a negative bias of 5-10% increasing towards the top of the profile. The apparent tension between the sonde and lidar comparison results in the polar regions can be traced to either poor co-location statistics and hence larger uncertainties on the bias estimates (Arctic), or to a known negative bias of the lidar data between 1991 and 1998 at Dumont d’Urville (Antarctica, [RD41]).

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above the 20 hPa level (~25 km) the half-IP-68 spread is 4-5% at the equator and increases gradually up to 8-10% at the poles. The comparison spread increases rapidly at lower altitudes, and reaches 40% and more around tropopause. The spread seen in the comparisons is a clearly larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ (not shown here) provided in the HALOE data files. The latter is only 1-3% at altitudes above the 50 hPa level (20 km) and increases to about 20-40% at the tropopause.

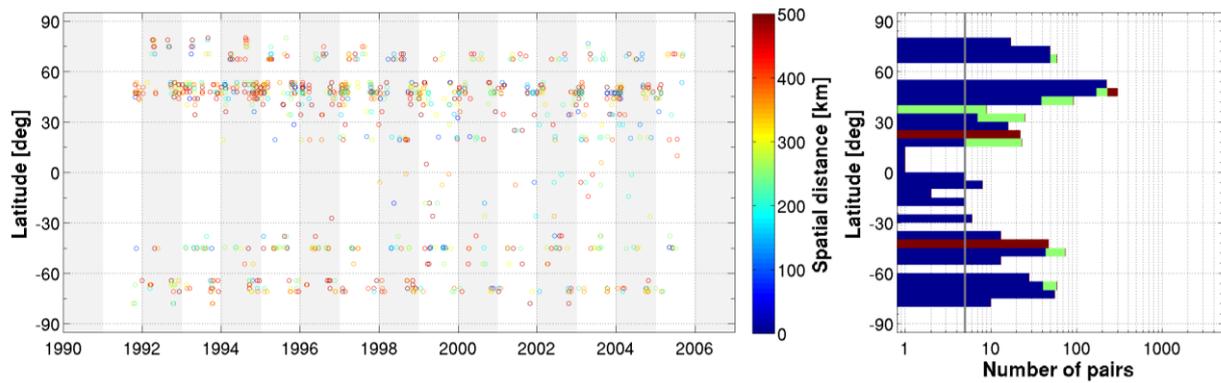


Figure 6.47 - (Left) Latitude–time distribution of co-locations between HALOE v19 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

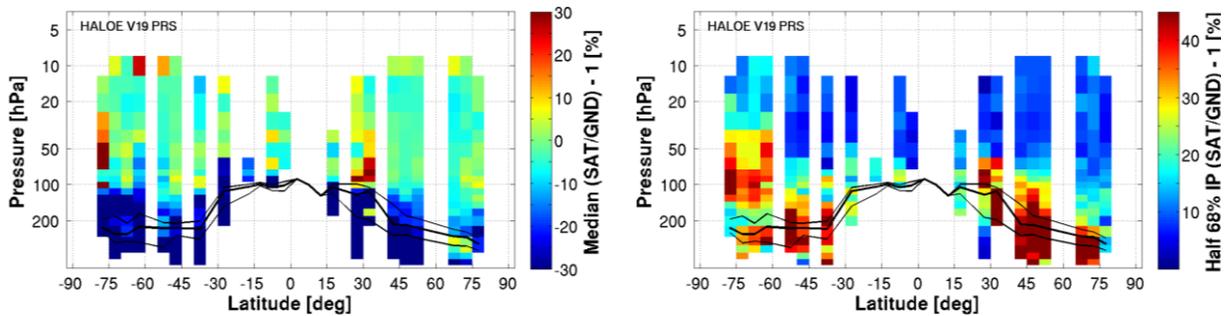


Figure 6.48 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between HALOE v19 ozone profile data and the global ozonesonde network, calculated over the entire HALOE time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

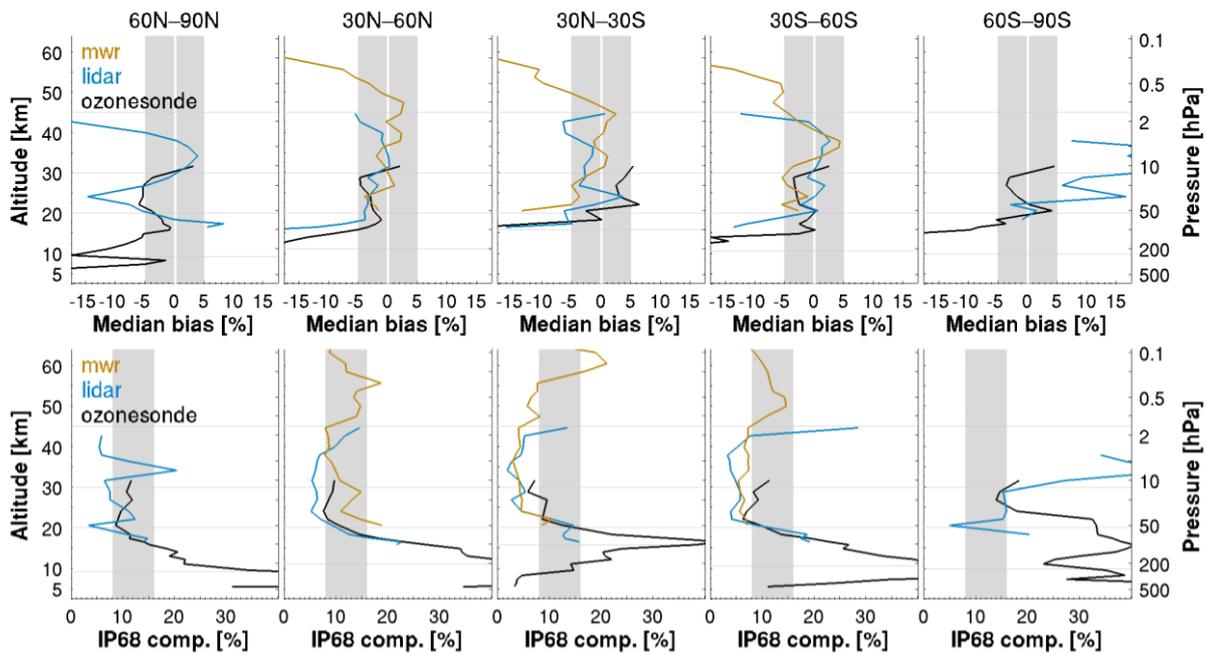


Figure 6.49 - (Top) Median bias between HALOE v19 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

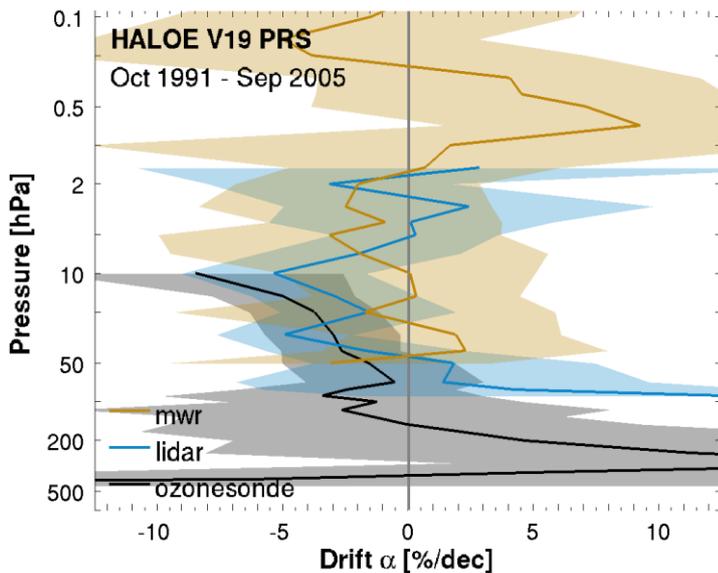


Figure 6.50 - Global average drift (in percent / decade) of HALOE v19 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 1991-2005 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.13.3 Long-term stability

Figure 6.50 shows the vertical structure of the decadal drift of HALOE v19 data relative to the ozonesonde, lidar and MWR networks. Drift estimates are negative between 200-2 hPa and significant between 10-40 hPa (~23-30 km) for both the sonde and the lidar comparisons. In the lower mesosphere values are positive but insignificant. The negative drift in the middle stratosphere amounts to 3-4% per decade. We therefore advice caution when using HALOE data for studies of long-term changes in ozone.



6.4.13.4 Dependence of data quality on other parameters

The bias, the comparison spread and the long-term stability of HALOE ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the HALOE HARMOZ data files and in the correlative data files are consistent.

6.4.13.5 Summary table of validation results

Table 6.24 - Summary of comparison results between HALOE v19 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	200-50 hPa	50-10 hPa	10-1 hPa	1-0.1 hPa
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-30, 0]	[-6, +2]	[-15, +3]	–
• Middle latitudes	[-30, -1]	[-5, 0]	[-4, +2]	[-15, +2]
• Tropics	[-30, 0]	[-4, +4]	[-6, +1]	[-15, +2]
Range comparison spread (half 68% interpercentile; %)				
• Polar	10/23–35	8–12 / 16–34	7–40	–
• Middle latitudes	12–40	5–10	5–8	8–16
• Tropics	15–40	4–14	3–5	5–16
Range long-term stability (drift; %/decade)				
• Ground network	[-2, +4]	[-6, -2]	[-5, +2]	[-5, +9]



6.4.13.6 Compliance with user requirements

From the results reported above it can be concluded that the HALOE v19 ozone profile data record is compliant with most of sampling and resolution requirements. Data quality meets the requirements on random uncertainty at altitudes above the 50 hPa level (~20 km). A significant negative drift of 3-4% per decade is seen between 10-40 hPa (~23-30 km) and has to be considered in long-term studies. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, compliance is flagged as not fulfilled in the Antarctic, but it is understood that this may not reflect the actual data quality.

Table 6.25 - Compliance of HALOE v19 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain [RD88]
Vertical resolution	< 1–3 km	2.3 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	10/1991 – 11/2005
Total uncertainty in height registration	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude, sunset/sunrise

Layer [hPa]	Lower stratosphere				Middle atmosphere				
	200-100	100-50	50-20	20-10	10-5	5-2	2-1	1-0.5	0.5-0.1
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic	(large natural variability)								
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic									
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.14 TIMED SABER 9.6 μm v2.0

Identification data record	
Observation principle	Infrared limb emission, along track scanning
Platform	TIMED, 74.0° inclination orbit
Responsible institute	Hampton University, University of Bremen – IUP
Contact person	James Russell III (james.russell@hamptonu.edu), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	01/2002 – 07/2021
• Latitude	90°S – 90°N
• Longitude	180°W – 180°E
• Vertical	ALT : 5 – 108 km (104 levels), PRS : 500 – 10 ⁻⁴ hPa (56 levels)
L1 processor and version	–
L2 processor and version	L2A v2.0 – 9.6 μm
Validated L2 file version	PRS : fv0005
Retrieval representation	O3 volume mixing ratio versus geometric pressure

6.4.14.1 Co-locations / Validation sample

Figure 6.51 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with SABER 9.6 μm v2.0 data. The sampling is very dense as SABER measures ~1400 profiles per day [RD74]. All latitude zones are covered and temporal sampling is very homogeneous. Slightly fewer co-locations are also seen in the last few years of the mission, due to the unavailability of publicly released correlative data.

6.4.14.2 Bias and spread

The median bias and the half 68% interpercentile of the relative difference between SABER 9.6 μm v2.0 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.52 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.53 shows the same information calculated in 30° latitude zones.

All ground-based comparisons show a similar pronounced vertical structure of SABER bias. Around 30-40 hPa (~21-24 km) SABER bias ranges from -5% to 0%. At higher altitudes, there is a clear overestimation which increases with altitude, reaching 5-10% (NH) to 10-15% (NH) at the stratopause. SABER overestimates ozone w.r.t. ozonesonde in the lowermost stratosphere by 5-15% around 50 hPa. This vertical structure is similar to that reported for an earlier version of the SABER data [RD74]. Opposed to the previous SABER data release (fv0003), the current file version (fv0005) does include measurements below the 50 hPa level. We recommend caution in using SABER ozone profile data in the UTLS and below due to a positive bias of at least 30%.

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. In the upper stratosphere, the half-IP-68 spread is about 6-8% at the equator and increases gradually towards the poles. Higher variability is seen in the mesosphere (~15%) and middle stratosphere (~12%). In the lower stratosphere and below, dispersion in the comparisons rapidly increases from 15% to more than 40%. Ex-ante uncertainties in the current SABER data release (fv0005) have decreased significantly with respect to the previous file version (fv0003). Median reported uncertainty drops from 16% to 5% between 10-50 hPa and from 9% to 1% at pressures below 2 hPa. These much smaller ex-ante uncertainties appear too optimistic, as the observed spread in the comparisons is much higher across the entire profile.

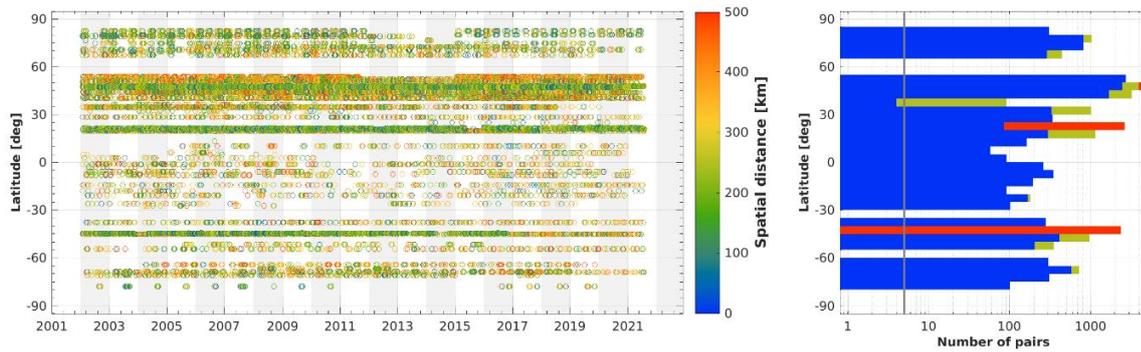


Figure 6.51 - (Left) Latitude–time distribution of co-locations between SABER 9.6 μm v2.0 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

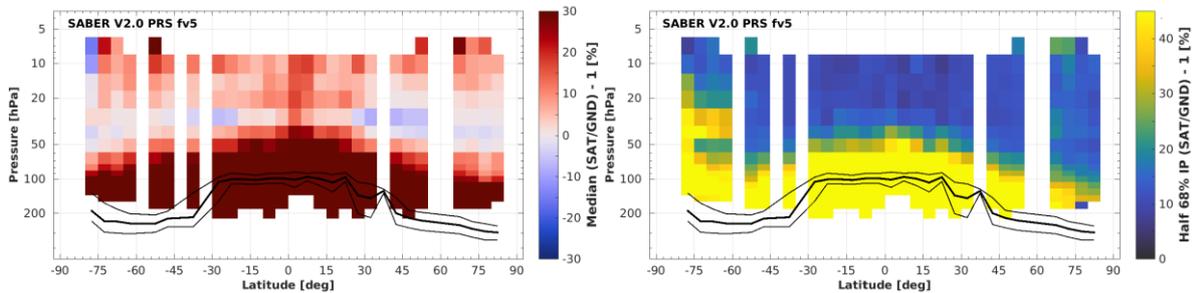


Figure 6.52 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between SABER 9.6 μm v2.0 ozone profile data and the global ozonesonde network, calculated over the entire SABER time period and in 5° bins. Black lines indicate the median (thick) and 1 σ spread (thin) of the tropopause altitude.

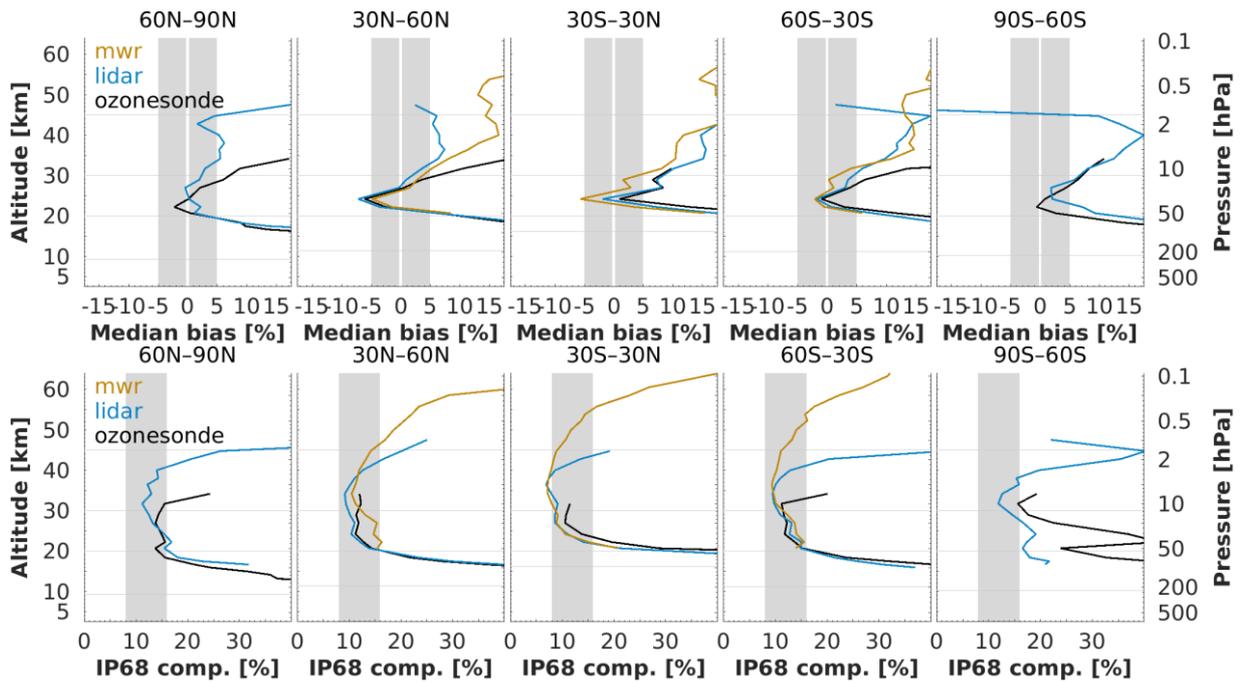


Figure 6.53 - (Top) Median bias between SABER 9.6µm v2.0 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

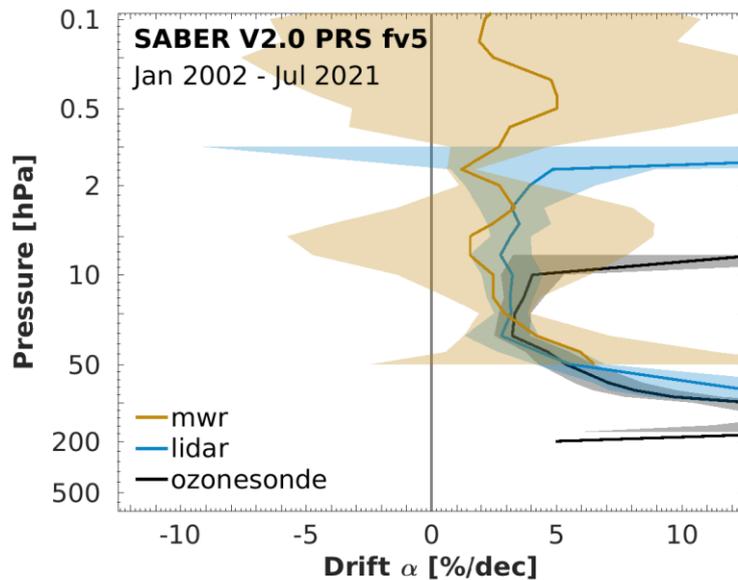


Figure 6.54 - Global average drift (in percent / decade) of SABER 9.6µm v2.0 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2002-2021 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.14.3 Long-term stability

Figure 6.54 shows the drift of SABER 9.6µm v2.0 ozone profile data with respect to co-located ground-based network data. We find a consistent picture that SABER ozone drifts towards higher values relative to all independent ground-based data records, across the entire profile. Our drift estimates are highly significant from the stratopause down to the lowest profile level. SABER drifts by more than +5-10% per decade below the 50 hPa level and by about +3% per decade at all higher altitudes. We recommend great caution when using these data for studies of long-term changes in ozone.



6.4.14.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of SABER 9.6 μ m v2.0 ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the SABER HARMOZ data files and in the correlative data files are consistent.

6.4.14.5 Summary table of validation results

Table 6.26 - Summary of comparison results between SABER 9.6 μ m v2.0 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	200-50 hPa	50-10 hPa	10-1 hPa	1-0.1 hPa
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	> +5	[-4, +6]	[0, +20]	–
• Middle latitudes	> +10	[-6, +7]	[+3, +12]	[+14, +18]
• Tropics	> +15	[+1, +9]	[+5, +15]	[+10, +18]
Range comparison spread (half 68% interpercentile; %)				
• Polar	16-40	13–15/30	13–40	–
• Middle latitudes	16-50	10–14	8–16	16–32
• Tropics	30-60	8–20	6–9	9–32
Range long-term stability (drift; %/decade)				
• Ground network	[+5, +15]	[+3, +5]	[+2, +3]	[+2, +4]



6.4.14.6 Compliance with user requirements

From the results reported above it can be concluded that the SABER 9.6 μm v2.0 ozone profile data record is compliant with most of sampling and resolution requirements. Data quality does not meet the requirements over most of the profile. Random uncertainty exceeds the 8% threshold nearly everywhere and there is solid evidence of a +3% per decade and more drift across the entire profile which will impact long-term studies.

Table 6.27 - Compliance of SABER 9.6 μm v2.0 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	500 km [RD74]
Vertical resolution	< 1–3 km	2 km [RD74]
Observation frequency	< 3 days	3 days
Time period	(1980-2010) – (2003-2010)	01/2002 – 07/2021
Total uncertainty in height registration	< \pm 500 m	unknown
Dependences	–	latitude, altitude

Layer [hPa]	Lower stratosphere				Middle atmosphere				
	200-100	100-50	50-20	20-10	10-5	5-2	2-1	1-0.5	0.5-0.1
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 									
Long-term stability									
User requirement	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.15 EOS-Aura MLS v4.2

Identification data record	
Observation principle	Millimetre wave limb emission
Platform	EOS-Aura, polar orbit, afternoon sun-synchronous precession
Responsible institutes	JPL NASA, University of Bremen – IUP
Contact persons	Lucien Froidevaux (Lucien.Froidevaux@jpl.nasa.gov), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	08/2004 – 02/2020
• Latitude	82°S – 82°N
• Longitude	180°W – 180°E
• Vertical	PRS : 500 – 0.02 hPa (40 levels)
L1 processor and version	–
L2 processor and version	v4.2
Validated L2 file version	PRS : fv0007
Retrieval representation	O3 volume mixing ratio versus pressure

6.4.15.1 Co-locations / Validation sample

Figure 6.55 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with Aura MLS v4.2 data. The sampling covers most latitude zones and is very homogeneous in time. Slightly fewer co-locations are also seen in the last few years of the mission, due to the unavailability of publicly released correlative data.

6.4.15.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between Aura MLS v4.2 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.56 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.57 shows the same information calculated in 30° latitude zones.

The vertical-meridian structure of the median bias follows that of the tropopause. Stratospheric and mesospheric bias is mostly less than 3-4% and has almost no vertical structure. Imprints of vertical oscillations are clearly visible in the UTLS, around 100 hPa (~15 km). This feature is known [RD63] but is somewhat washed out in the Ozone_cci+ product due to the different vertical grid than the original Aura MLS data.

The meridian structure of the comparison spread $s_{\Delta x}$ follows that of the tropopause. Above 30-40 hPa (~21-24 km) the half-IP-68 spread is 3-4% at the equator and increases gradually up to 6-8% at the poles. At lower altitudes, the comparison spread increases rapidly. Around the tropopause the spread is 25% in the tropics and 35-40% at higher latitudes. The spread seen in the comparisons is generally larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ (not shown here) provided in the Aura MLS data files (2-3% above 50 hPa/20 km).

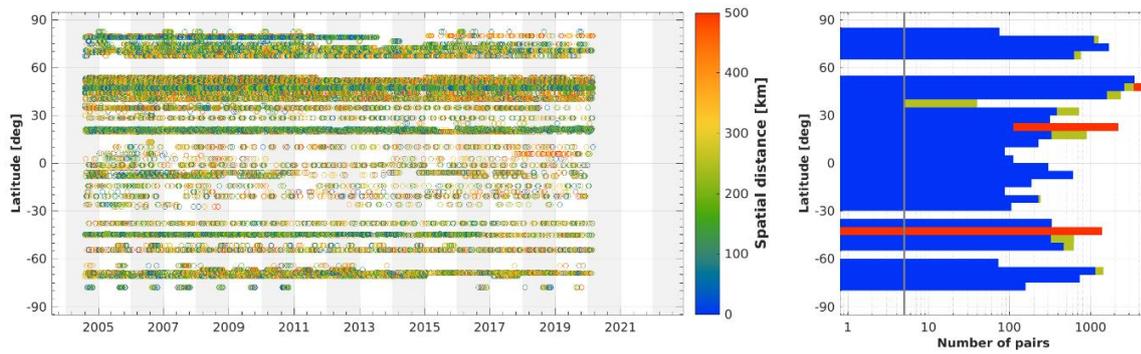


Figure 6.55 - (Left) Latitude–time distribution of co-locations between Aura MLS v4.2 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

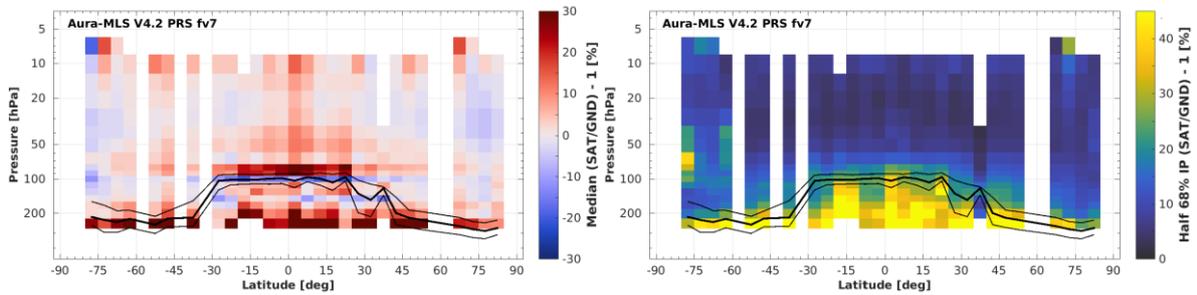


Figure 6.56 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between Aura MLS v4.2 ozone profile data and the global ozonesonde network, calculated over the entire Aura MLS time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

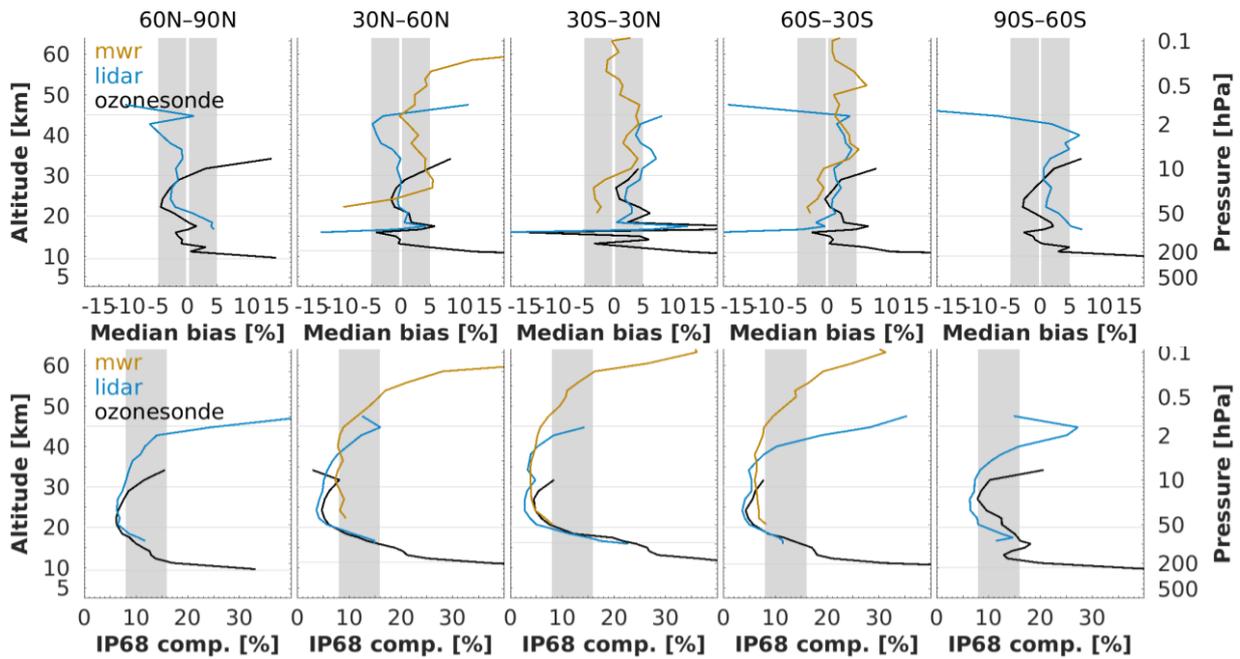


Figure 6.57 - (Top) Median bias between Aura MLS v4.2 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

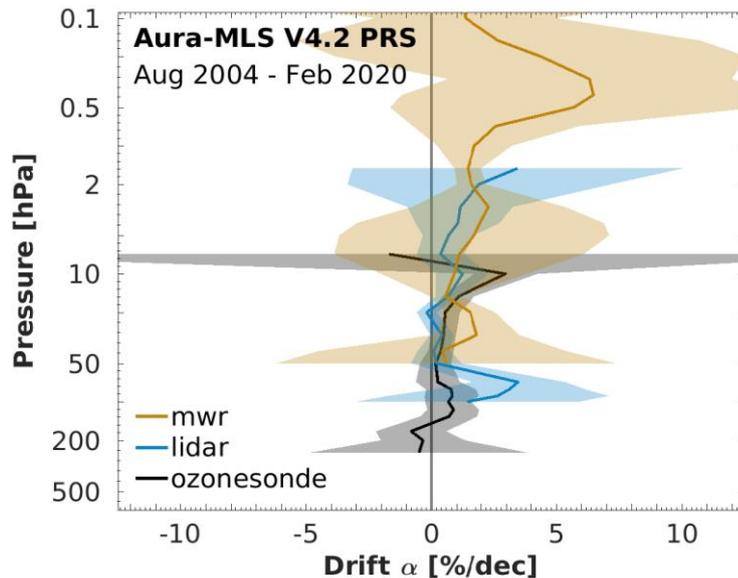


Figure 6.58 - Global average drift (in percent / decade) of Aura MLS v4.2 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2004-2020 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.15.3 Long-term stability

Figure 6.58 shows the drift of Aura MLS v4.2 ozone profile data with respect to the co-located ozonesonde, lidar and MWR network data. Drift estimates are less than 1% per decade between 200-3 hPa (~10-42 km) and less than 2% per decade between 3-0.7 hPa (~42-48 km). Confidence in the +6% per decade peak around 0.4 hPa is low. Overall, results from all ground-based techniques are in good agreement in regions where they are most sensitive. There is therefore substantial support for a drift-free Aura MLS data record (at the



level of about 1% per decade) across the entire stratosphere and possibly the good stability extends into the mesosphere as well.

6.4.15.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of Aura MLS v4.2 ozone are very similar in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the Aura MLS HARMOZ data files and in the correlative data files are consistent.

An important note to users of Aura MLS data is that two optional parameters in the HARMOZ Aura MLS data files (*altitude_mls* and *geopotential_height_mls*) were taken from the original Aura MLS Level-2 data and should not be used. Both variables are affected by considerable uncertainties in the absolute pointing of the instrument, especially in the first four years of the mission [RD63]. The *altitude* profiles provided in HARMOZ data were obtained by interpolating co-located ERA5 altitude and pressure profiles to the reported MLS pressure levels before regridding to the HARMOZ vertical grid.

6.4.15.5 Summary table of validation results

Table 6.28 - Summary of comparison results between Aura MLS v4.2 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	200-50 hPa	50-10 hPa	10-1 hPa	1-0.1 hPa
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-3, +15]	[-4, +2]	[-5, +5]	–
• Middle latitudes	[-5, +15]	[-1, +2]	[-5, +4]	[0, +9]
• Tropics	[-15, +15]	[+1, +5]	[+3, +6]	[-1, +4]
Range comparison spread (half 68% interpercentile; %)				
• Polar	6–25 / 14–20	6–8	8–32	–
• Middle latitudes	6–32	3–6	4–11	8–28
• Tropics	8–32	3–6	4–6	6–26
Range long-term stability (drift; %/decade)				
• Ground network	[-1, +1]	[0, +1]	[+1, +2]	[+2, +6]



6.4.15.6 Compliance with user requirements

From the results reported above it can be concluded that the Aura MLS v4.2 ozone profile data record is nearly compliant with all sampling and resolution requirements. Requirements on random uncertainty are met at altitudes above the 50 hPa level (~20 km). Also decadal stability is compliant over the entire stratosphere and mesosphere. Assessing compliance of random uncertainty in polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed comparison spread. As a result, random uncertainty compliance is flagged as not fulfilled in the polar upper stratosphere, but it is understood that this may not reflect the actual data quality.

Table 6.29 - Compliance of Aura MLS v4.2 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	200–500 km [RD63]
Vertical resolution	< 1–3 km	2.5–5 km [RD63]
Observation frequency	< 3 days	Daily global coverage [RD93]
Time period	(1980-2010) – (2003-2010)	08/2004 – 10/2017
Total uncertainty in height registration	< ± 500 m	unknown
Dependences	–	latitude, altitude

Layer [hPa]	Lower stratosphere				Middle atmosphere				
	200-100	100-50	50-20	20-10	10-5	5-2	2-1	1-0.5	0.5-0.1
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
• Arctic					(large natural variability)				
• Mid NH									
• Tropics									
• Mid SH									
• Antarctic					(large natural variability)				
Long-term stability									
User requirement	< 1-3% per decade								
• Ground network									



6.4.16 SPOT-4 POAM III v4

Identification data record	
Observation principle	UV-visible solar occultation
Platform	SPOT-4, 98.7° inclination orbit
Responsible institute	Naval Research Lab – NRL, University of Bremen – IUP
Contact person	Richard Bevilacqua (bevilacqua@nrl.navy.mil), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	04/1998 – 12/2005
• Latitude	88°S–63°S (sunset), 55°N–71°N (sunrise)
• Longitude	180°W – 180°E
• Vertical	ALT : 5 – 60 km (56 levels)
L1 processor and version	–
L2 processor and version	v4.0
Validated L2 file version	ALT : fv0002
Retrieval representation	O3 number density versus geometric altitude

6.4.16.1 Co-locations / Validation sample

Figure 6.31 shows the latitude–time distribution of the ozonesonde and lidar measurements co-locating with the POAM III v4 dataset. Due to the orbital inclination of the satellite and its solar occultation viewing geometry, comparisons are mostly found at high latitudes. Co-locations are noted at two sites at Northern mid-latitudes but we deem that results are likely not representative outside of the polar regions. No co-locations with MWR measurements were found. The temporal sampling is fairly continuous given annual interruptions during polar winter. In addition, the type of occultation differs between the Northern (sunrise) and Southern (sunset) Hemispheres, which should be considered in the (lidar) comparisons above ~35 km where the magnitude of the diurnal cycle in ozone becomes stronger.

6.4.16.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between POAM III v4 and GAW/NDACC/SHADOZ ozonesondes and NDACC lidars are shown in the figures below. Figure 6.59 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.60 shows the same information calculated in 30° latitude zones.

POAM III underestimates ground-based data by up to 5% between 15-30 km at high latitudes. At lower and higher altitudes, there is a positive bias with respect to sonde and/or lidar, which reaches 10-15% close to the tropopause and 5-10% in the Arctic upper stratosphere. The apparent tension between the sonde and lidar comparison results in the polar regions can be traced to a known negative bias of the lidar data between 1991 and 1998 at Dumont d’Urville (Antarctica, [RD41]).

The comparison spread $s_{\Delta x}$ in the Arctic is 8-10% between 20-30 km and increases to 30-40% at the tropopause and at 40 km altitude. In the Antarctic the variability is somewhat larger due to the presence of the ozone hole. The spread seen in the comparisons is slightly larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ provided in the POAM III record (not shown here). The latter is of the order of 4-7% between 20-30 km and increases up to 20-40% in the UTLS and upper troposphere.

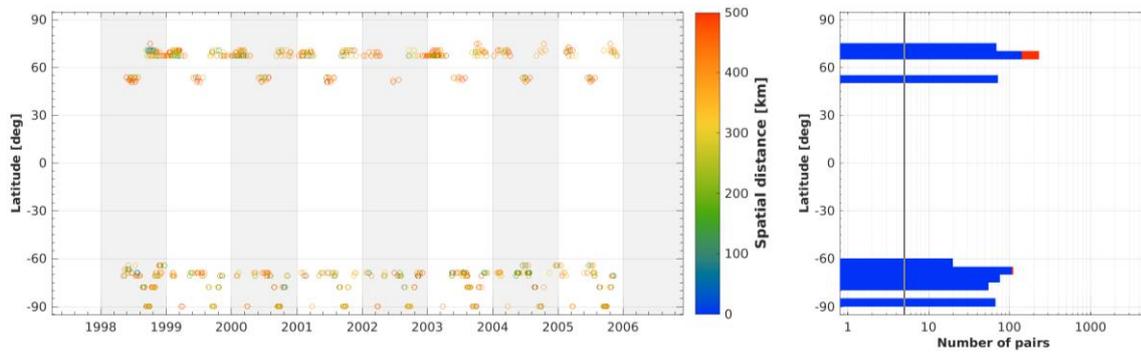


Figure 6.59 - (Left) Latitude–time distribution of co-locations between POAM III v4 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde and NDACC stratospheric ozone lidar). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue) and lidar (red).

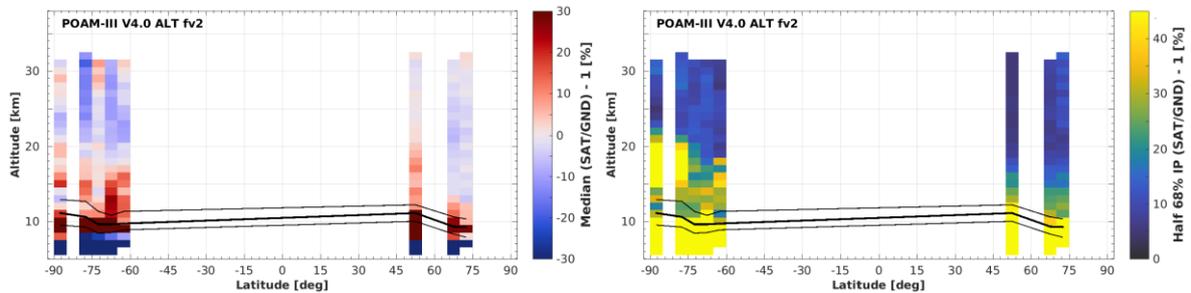


Figure 6.60 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between POAM III v4 ozone profile data and the global ozonesonde network, calculated over the entire POAM III time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude. Results outside the polar regions are likely not representative.

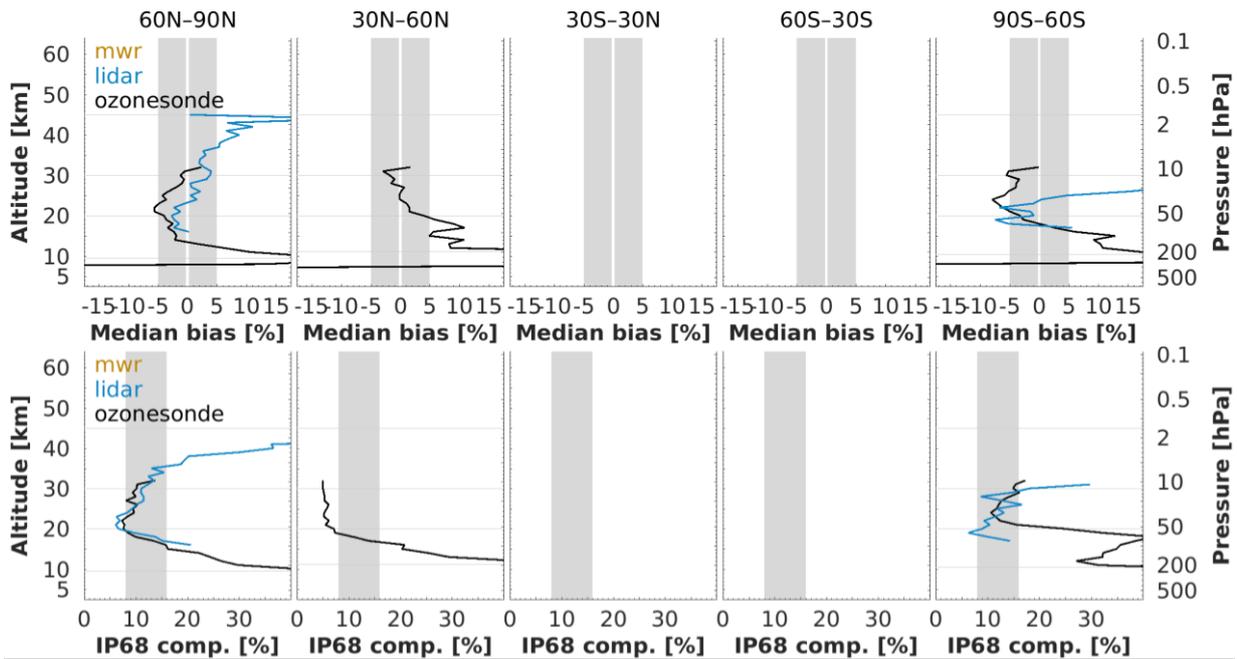


Figure 6.61 - (Top) Median bias between POAM III v4 ozone profile data and ozonesonde (black) and lidar (blue) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample. Results outside the polar regions are likely not representative.

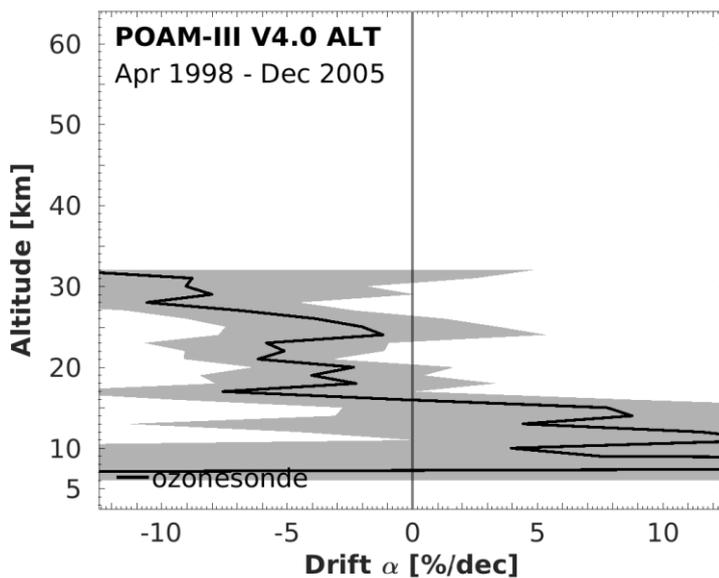


Figure 6.62 - Global average drift (in percent / decade) of POAM III v4 ozone profile data with respect to co-located ozonesonde (black) network data, calculated over the 1998-2005 time period. The shaded region represents 2σ uncertainty on the average decadal drift.

6.4.16.3 Long-term stability

Figure 6.34 shows the vertical structure of the decadal drift of POAM III v4 data relative to the ozonesonde network. There are not enough co-locations to obtain meaningful estimates of drift relative to lidar measurements. Co-locations with ozonesonde are available at just a handful of sites due to POAM III's limited spatial coverage. Estimates of network-averaged drift will therefore be more prone to inhomogeneities (spatial and temporal) in the ozonesonde records than those for other satellite sensors. The drift uncertainty, estimated at 2-4% per decade between 20-30 km (1σ), is likely smaller than in reality. For this reason, we



have low confidence in the 5-10% per decade negative drift between 20-30 km. Nonetheless, we recommend users of POAM III data to verify their stability as our analysis can not exclude the presence of drift.

6.4.16.4 Dependence of data quality on other parameters

Bias and comparison spread of POAM III ozone are in excellent agreement in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the POAM III HARMOZ data files and in the correlative data files are consistent.

6.4.16.5 Summary table of validation results

Table 6.30 - Summary of comparison results between POAM III v4 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-3, +18]	[-7, +1]	[+2, +8]	–
• Middle latitudes	results not representative			
• Tropics	no data			
Range comparison spread (half 68% interpercentile; %)				
• Polar	7–40	7–16	1–32	–
• Middle latitudes	results not representative			
• Tropics	no data			
Range long-term stability (drift; %/decade)				
• Ground network	[-5, +7]	[-9, -1]	–	–



6.4.16.6 Compliance with user requirements

Our ground-based assessment of the quality of the POAM III v4 ozone profile data record is restricted to high latitudes. The data record meets the user’s uncertainty requirements only in the Arctic lowermost stratosphere. However, assessing random uncertainty in the polar regions proves challenging due to the unquantified but likely considerable contribution by natural variability to the observed dispersion in the comparisons. Furthermore, the ground-based analysis is unable to test compliance of POAM III stability at the 1-3% per decade level, mainly due to the scarcity of comparison data.

Table 6.31 - Compliance of POAM III v4 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain [RD88]
Vertical resolution	< 1–3 km	1-2 km
Observation frequency	< 3 days	not compliant
Time period	(1980-2010) – (2003-2010)	04/1998 – 12/2005
Total uncertainty in height attribution	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 									
Long-term stability									
User requirement	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.17 Meteor-3M SAGE III v4

Identification data record	
Observation principle	UV-visible solar occultation
Platform	Meteor-3M, 99.7° inclination orbit
Responsible institutes	NASA Langley Research Center, University of Bremen – IUP
Contact persons	Robert Damadeo (robert.damadeo@nasa.gov), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	06/2002 – 11/2005
• Latitude	50°S–30°S, 50°N–80°N
• Longitude	180°W – 180°E
• Vertical	ALT : 5 – 65 km (61 levels)
L1 processor and version	–
L2 processor and version	v4
Validated L2 file version	ALT : fv0001
Retrieval representation	O3 number density versus geometric altitude

6.4.17.1 Co-locations / Validation sample

Figure 6.63 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with SAGE III/M3M v4 data. The sampling pattern is quite peculiar due to the orbit of the satellite. There is no homogeneous sampling in the time domain and comparisons are only available at mid-latitudes and in the Arctic.

6.4.17.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between SAGE III/M3M v4 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.64 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.65 shows the same information calculated in 30° latitude zones.

The SAGE III/M3M bias is less than $\pm 5\%$ in the middle and lower stratosphere, underestimating ground-based data in the Arctic and overestimating at mid-latitudes. Larger positive biases of 5-15% are noticed in the uppermost stratosphere, the lower mesosphere and in the UT/LS region. Between 20-40 km, the half-IP-68 spread lies between 3-8% at mid-latitudes and between 8-16% in the Arctic. The comparison spread increases rapidly at lower altitudes and reaches 40% around the tropopause. The spread seen in the comparisons is clearly larger than the ex-ante random uncertainty $\sigma_{\text{ex-ante}}$ (not shown here) provided in the SAGE III/M3M data files. The latter is less than 1-2% above 20 km and increases to about 25% at the tropopause.

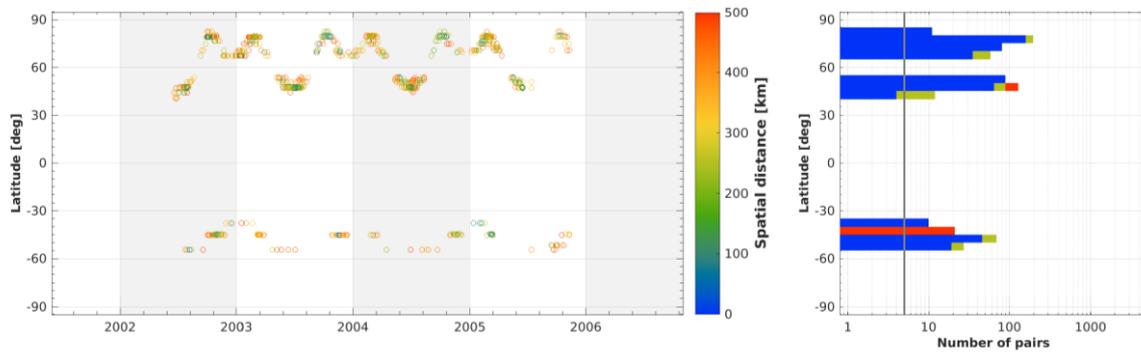


Figure 6.63 - (Left) Latitude–time distribution of co-locations between SAGE III/M3M v4 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

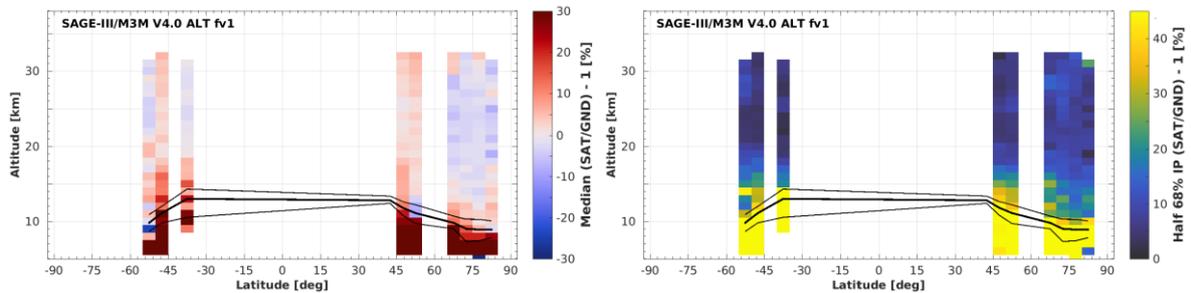


Figure 6.64 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between SAGE III/M3M v4 ozone profile data and the global ozonesonde network, calculated over the entire SAGE III/M3M time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

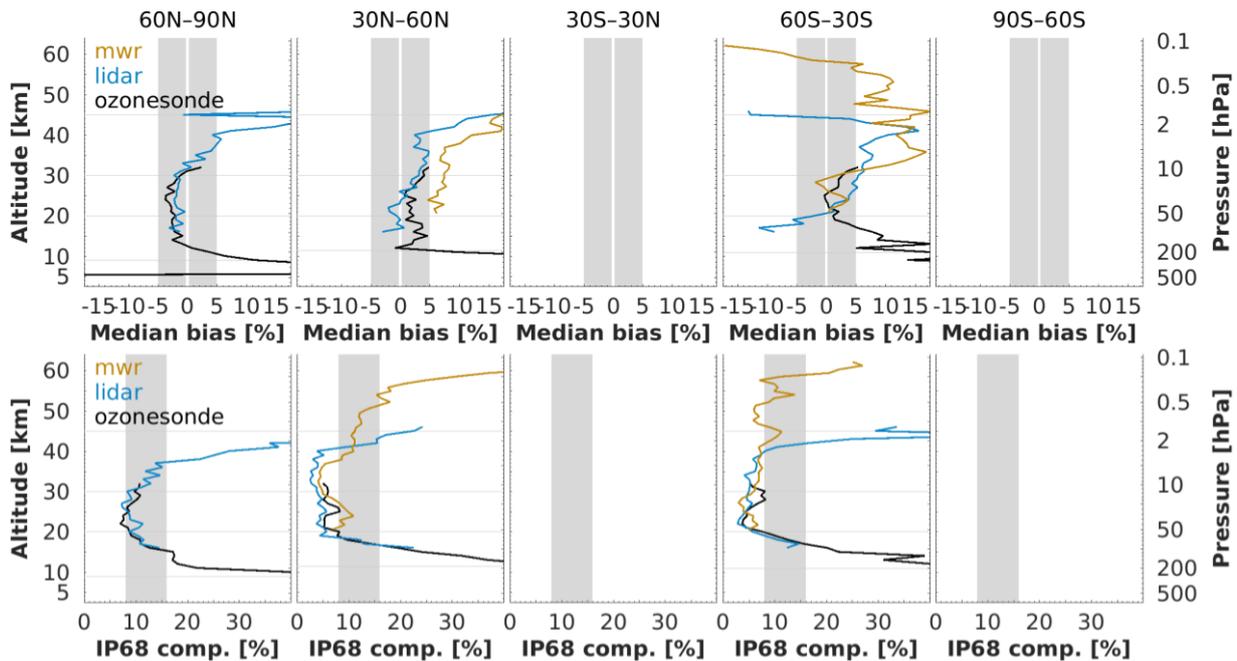


Figure 6.65 - (Top) Median bias between SAGE III/M3M v4 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

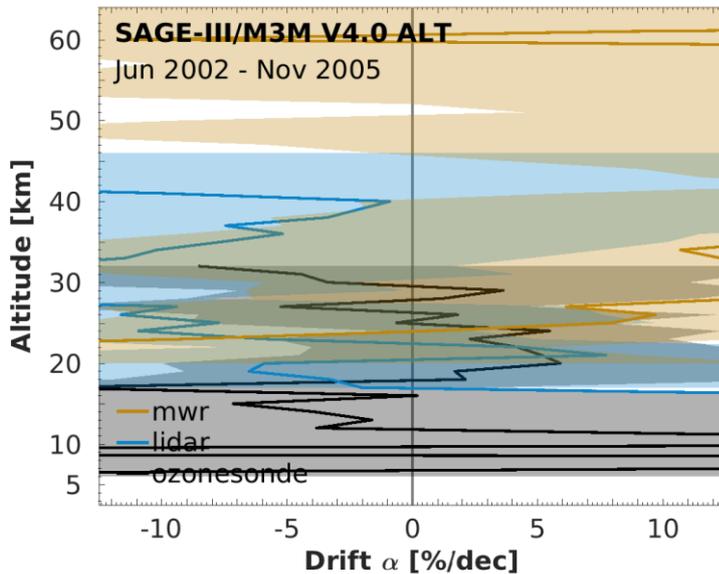


Figure 6.66 - Global average drift (in percent / decade) of SAGE III/M3M v4 ozone profile data with respect to co-located ozonesonde (black), lidar (blue) and MWR (orange) network data, calculated over the 2002-2005 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.17.3 Long-term stability

Figure 6.66 shows the vertical structure of the decadal drift of SAGE III/M3M v4 data relative to the ozonesonde, lidar and MWR networks. Due to its short data record SAGE III/M3M drift cannot be well constrained. The uncertainty (1 σ) of the estimates is at best 5-10% per decade across the stratosphere. Hence, we can only conclude that SAGE III/M3M drift, if any, is not worse than 10% per decade.



6.4.17.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of SAGE III/M3M ozone are in excellent agreement in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the SAGE III/M3M HARMOZ data files and in the correlative data files are consistent.

6.4.17.5 Summary table of validation results

Table 6.32 - Summary of comparison results between SAGE III/M3M v4 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar & MWR	MWR
Range systematic uncertainty (median bias; %)				
• Arctic	[-2, +10]	[-4, -2]	[-2, +9]	–
• Middle latitudes	[+1, +13]	[0, +4]	[+3, +15]	[-10, +20]
• Tropics	no data			
Range comparison spread (half 68% interpercentile; %)				
• Arctic	9–40	8–10	10–40	–
• Middle latitudes	9–40	4–6	5–10	8–35
• Tropics	no data			
Range long-term stability (drift; %/decade)				
• Ground network	data record too short			



6.4.17.6 Compliance with user requirements

Verifying the compliance of the short SAGE III/M3M data record with user requirements is very hard. What is already clear is that an UV-visible occultation mission, like SAGE, is unable to meet the temporal sampling requirements but that it does have the required vertical resolution and accuracy in altitude registration. Besides this, the random uncertainty requirements between 20-40 km are likely to be met at middle latitudes and in the lowermost stratosphere of the Arctic. Verification of compliance with user requirements for long-term stability is not possible.

Table 6.33 - Compliance of SAGE III/M3M v4 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain [RD88]
Vertical resolution	< 1–3 km	0.5 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	06/2002 – 11/2005
Total uncertainty in height registration	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude, sunset/sunrise

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %				< 8%				
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 									
Long-term stability									
User requirement	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.18 ISS SAGE III v5.1

Identification data record	
Observation principle	UV-visible solar occultation*
Platform	ISS, 51.6° inclination orbit
Responsible institutes	NASA Langley Research Center, University of Bremen – IUP
Contact persons	David E. Flittner (david.e.flittner@nasa.gov), Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)
Coverage	
• Time	06/2017 – 11/2019
• Latitude	70°S – 70°N
• Longitude	180°W – 180°E
• Vertical	ALT : 5 – 65 km (61 levels)
L1 processor and version	–
L2 processor and version	v5.1
Validated L2 file version	ALT : fv0002
Retrieval representation	O3 number density versus geometric altitude

* Retrievals from lunar occultation data are released by NASA, but not used by the Ozone_cci project.

6.4.18.1 Co-locations / Validation sample

Figure 6.67 shows the latitude–time distribution of the ozonesonde, lidar and MWR measurements co-locating with SAGE III/ISS v5.1 data. The number of co-locations is quite low due to the short SAGE III/ISS data record. At least five profile matches are required to obtain meaningful validation results, which is the case at just a handful of sites. The aggregation of the comparison sample of individual stations in broad latitude bands allows to improve constraints on SAGE III/ISS data quality to some extent. However, it must be kept in mind that sampling uncertainty and spatial inhomogeneities in ground-based network data may contribute considerably to the observed differences. The region with best robustness of the results is at Northern mid-latitudes, elsewhere additional data will be very helpful in future analyses.

6.4.18.2 Bias and spread

The median bias and half the 68% interpercentile of the relative difference between SAGE III/ISS v5.1 and GAW/NDACC/SHADOZ ozonesondes, NDACC lidars or NDACC MWRs are shown in the figures below. Figure 6.68 details the vertical and meridian dependence of the bias and comparison spread at 5° latitude resolution, while Figure 6.69 shows the same information calculated in 30° latitude zones.

SAGE III/ISS generally overestimates ozone with respect to observations by the different ground-based networks. The positive bias is 2-4% between 20-40 km altitude. Below 20 km, ozonesonde comparisons indicate larger positive biases of SAGE III/ISS at all latitude bands. Results above 40 km have very low confidence at the moment. Our conclusions are in line with earlier published work [RD65,RD92].

The comparison spread $s_{\Delta x}$ between 25-40 km lies between 5-8%. A slow increase (up to 12% at the stratopause) is noted at higher altitudes and a rapid increase below 20 km (up to 45% at the tropopause). Similar values are found for SAGE II (Figure 6.45) and SAGE III/M3M (Figure 6.65). The spread seen in the comparisons is clearly larger than the ex-ante random uncertainty $s_{\text{ex-ante}}$ (not shown here) provided in the SAGE II data files. The latter is only 0.5-1.5% above 20 km and increases to about 15 % at the tropopause.

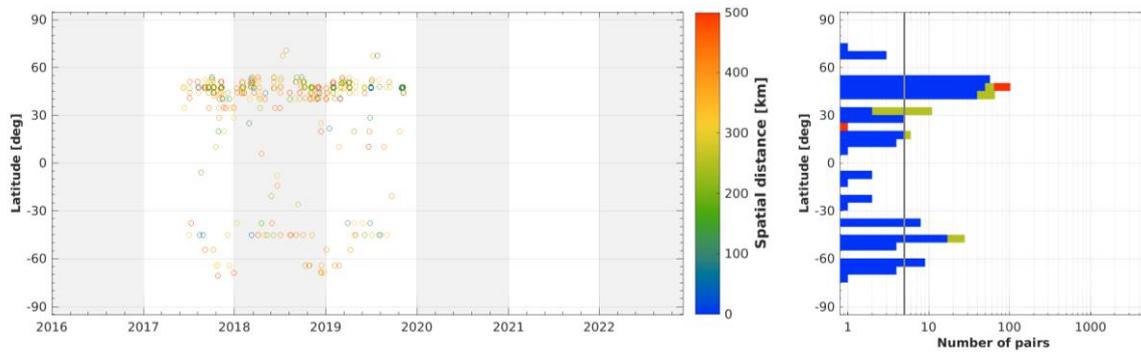


Figure 6.67 - (Left) Latitude–time distribution of co-locations between SAGE III/ISS v5.1 ozone profiles and ground-based measurements (GAW/NDACC/SHADOZ ozonesonde, NDACC stratospheric ozone lidar and NDACC MWR). The colour code indicates the spatial distance of each satellite/ground-based pair. (Right) Number of co-located pairs per 5° latitude band for ozonesonde (blue), lidar (green) and MWR (red).

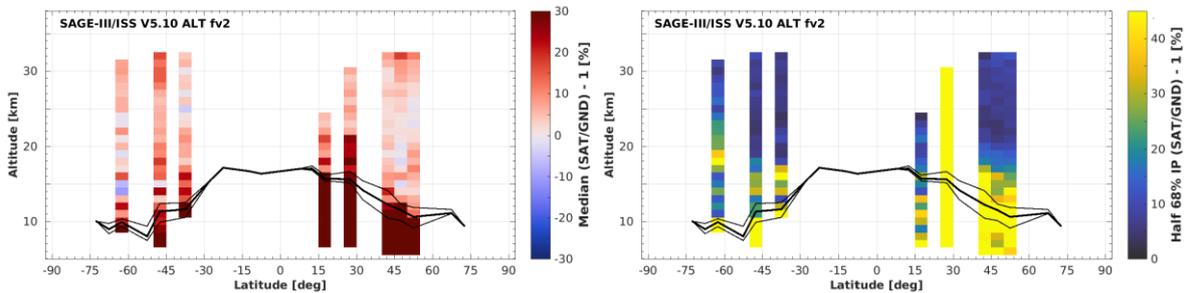


Figure 6.68 - Altitude–latitude cross-section of the median percent bias (left) and of the half IP-68 spread (right) between SAGE III/ISS v5.1 ozone profile data and the global ozonesonde network, calculated over the entire SAGE II time period and in 5° bins. Black lines indicate the median (thick) and 1σ spread (thin) of the tropopause altitude.

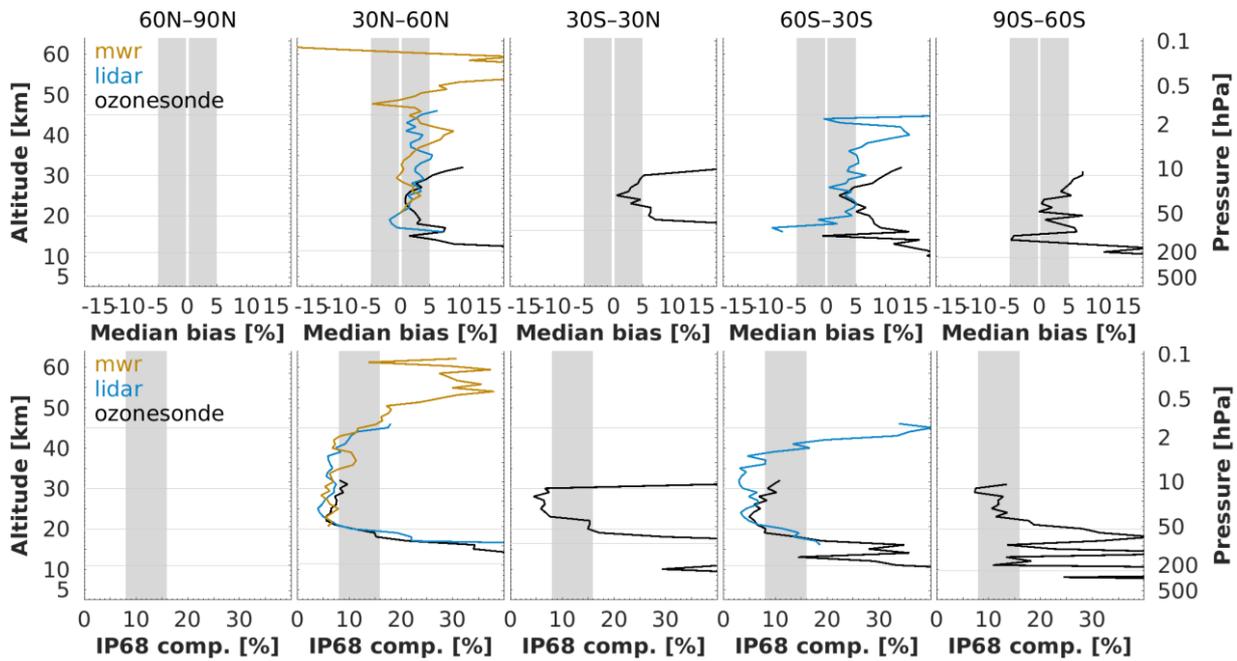


Figure 6.69 - (Top) Median bias between SAGE III/ISS v5.1 ozone profile data and ozonesonde (black), lidar (blue) and MWR (orange) data, by 30° latitude. (Bottom) Same, but for the half IP-68 spread. The lowest horizontal line indicates the median tropopause altitude over the co-location sample.

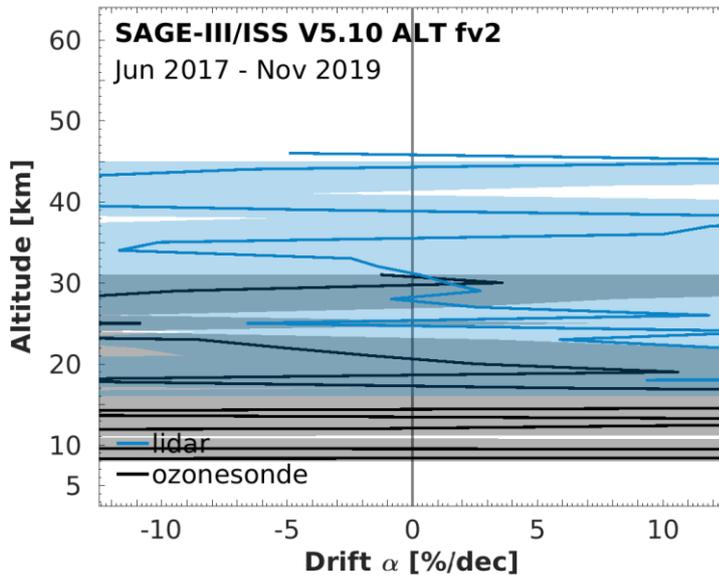


Figure 6.70 - Global average drift (in percent / decade) of SAGE III/ISS v5.1 ozone profile data with respect to co-located ozonesonde (black) and lidar (blue) network data, calculated over the 2017-2019 time period. The shaded region represents 2 σ uncertainty on the average decadal drift.

6.4.18.3 Long-term stability

Figure 6.70 shows the vertical structure of the drift of SAGE III/ISS v5.1 data relative to the ozonesonde and lidar networks. At the moment, SAGE III/ISS drift is very poorly constrained due to the short data record and the sparse set of co-located profiles. Uncertainty of drift estimates (1 σ) is 8-12% per decade between 20-30 km and larger at other altitudes. Longer time series are needed to verify the stability of the data record.



6.4.18.4 Dependence of data quality on other parameters

Bias, comparison spread and long-term stability of SAGE III/ISS ozone are in excellent agreement in four different profile representations. This indicates that the auxiliary pressure and temperature profiles included in the SAGE III / ISS HARMOZ data files and in the correlative data files are consistent.

6.4.18.5 Summary table of validation results

Table 6.34 - Summary of comparison results between SAGE III/ISS v5.1 and ground-based reference data. Values refer to the range of three data quality indicators in three zonal regions and four layers.

Layer	10-20 km	20-30 km	30-45 km	45-60 km
Reference data	sonde	sonde & lidar	lidar	MWR
Range systematic uncertainty (median bias; %)				
• Polar	[-5, +15]	[0, +6]	–	–
• Middle latitudes	[+3, +15]	[+1, +5]	[+1, +5]	–
• Tropics	> +6	[0, +6]	–	–
Range comparison spread (half 68% interpercentile; %)				
• Polar	15–40	8–30	–	–
• Middle latitudes	10–40	5–8	6–12	–
• Tropics	16–45	6–13	–	–
Range long-term stability (drift; %/decade)				
• Ground network	data record too short			



6.4.18.6 Compliance with user requirements

Verifying the compliance of the short SAGE III/ISS data record with user requirements is very hard at this point. For now, we report provisional results which will need to be consolidated after analysis of longer time series. What is already clear is that an UV-visible occultation mission, like SAGE, is unable to meet the temporal sampling requirements but that it does have the required vertical resolution and accuracy in altitude registration. Besides this, the random uncertainty requirements between 20-40 km are likely to be met at middle and low latitudes. Verification of long-term stability will require several more years of data.

Table 6.35 - Compliance of SAGE III/ISS v5.1 with user requirements (URD v3.1).

	Requirement	Compliance / evaluation
Horizontal resolution	< 100–300 km	uncertain [RD88]
Vertical resolution	< 1–3 km	1 km
Observation frequency	< 3 days	not compliant, ~30 solar occultation profiles per day
Time period	(1980-2010) – (2003-2010)	06/2017 – 11/2019
Total uncertainty in height registration	< ± 500 m	likely compliant (solar occultation)
Dependences	–	latitude, altitude, sunset/sunrise

Layer [km]	Lower stratosphere				Middle atmosphere				
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-60
Uncertainty including only random component									
User requirement	< 8-16 %		< 8%						
<ul style="list-style-type: none"> • Arctic • Mid NH • Tropics • Mid SH • Antarctic 									
Long-term stability									
User requirement	< 1-3% per decade								
<ul style="list-style-type: none"> • Ground network 									



6.4.19 Summary Graphs

The following figures give a global overview of the systematic uncertainty (Figure 6.71), random uncertainty (Figure 6.72) and decadal stability (Figure 6.73) of each Level-2 HARMOZ ozone profile data record, in five latitude zones. The last two viewgraphs indicate the user requirements of URD v3.1 [RD8]. MIPAS results relate to the 2005-2012 time period.

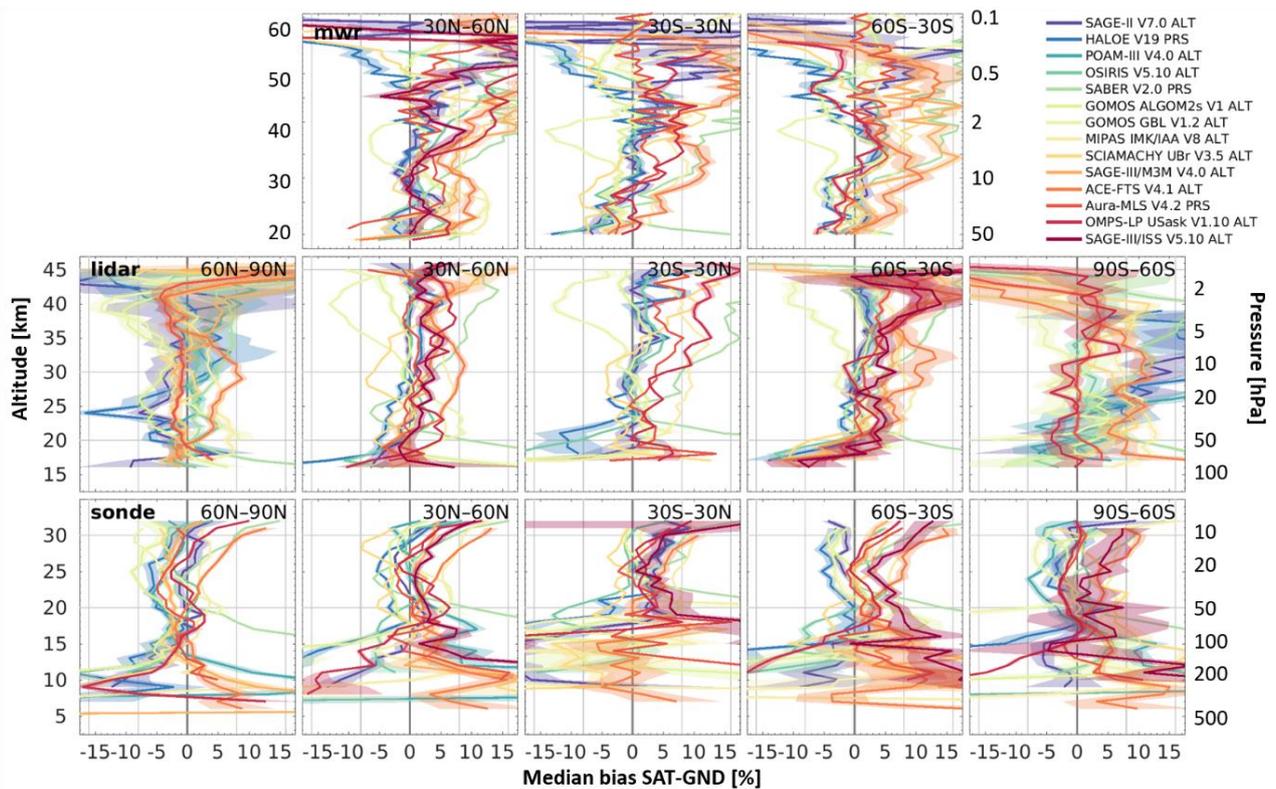


Figure 6.71 - Median bias of the Level-2 CRDP limb ozone profiles as a function of altitude/pressure and latitude (five zones from left to right). Comparisons are based on GAW/NDACC/SHADOZ ozonesonde (bottom), NDACC lidar (centre) and NDACC microwave radiometer (top) network data as reference. MIPAS results relate to the 2005-2012 time period.

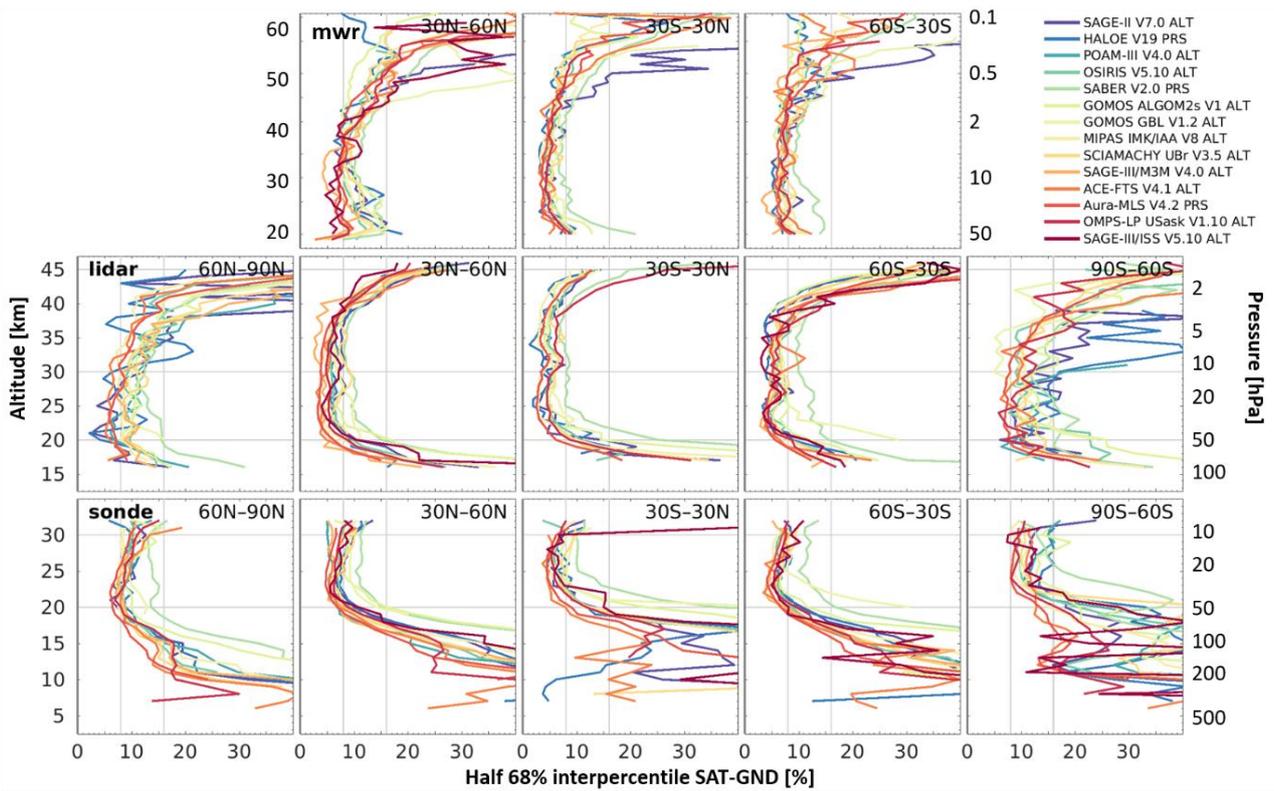


Figure 6.72 - Half width of the 68% interpercentile of the ground-based comparisons of Level-2 CRDP limb ozone profiles as a function of altitude/pressure and latitude (five zones from left to right). User requirement thresholds are indicated by thin vertical grey lines (URD v3.1). Comparisons are based on GAW/NDACC/SHADOZ ozonesonde (bottom), NDACC lidar (centre) and NDACC microwave radiometer (top) network data as reference. MIPAS results relate to the 2005-2012 time period.

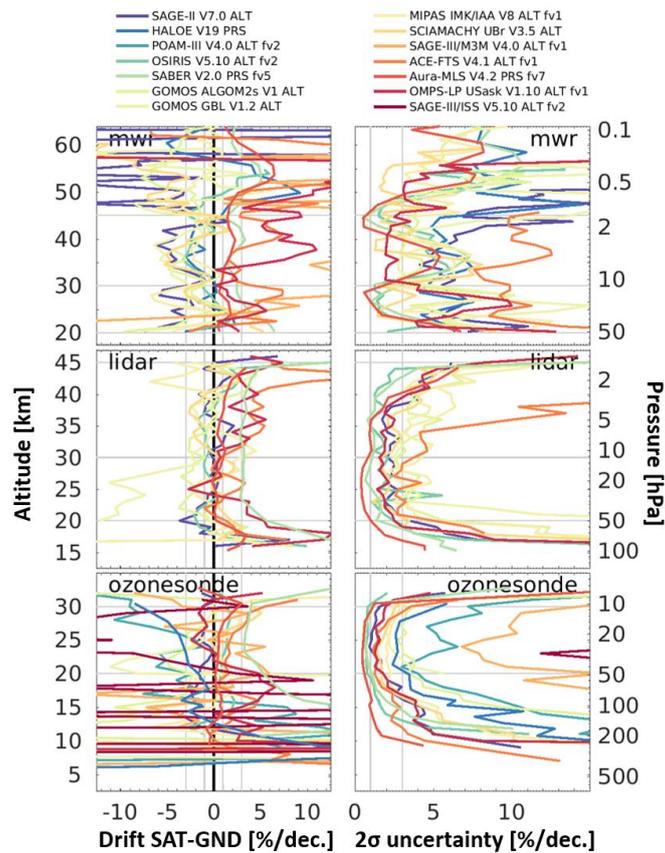


Figure 6.73 - Decadal stability (left) and its 2σ uncertainty (right) of CRDP limb ozone profile data as a function of altitude/pressure. User requirement thresholds are indicated by thin vertical grey lines (URD v3.1). Decadal stability is estimated as the drift between satellite data and reference measurements (GAW/NDACC/SHADOZ ozonesonde (bottom), NDACC lidar (centre) and NDACC MWR (top) network data). Time periods and nominal profile representation differ by satellite record. MIPAS results relate to the 2005-2012 time period.



6.5 Level-3 limb profile products

6.5.1 Evaluation method

The evaluation of gridded limb products (Level-3) differs considerably from the validation of single profiles (Level-2, Section 6.4.1). In a first step, the ozonesonde (and satellite) data are reshaped to comparable formats (Section 6.5.1.2). Then, quality indicators are derived from the comparison of the different incarnations of these data sets (Section 6.5.1.3). We start this section by motivating the need for such an adapted approach.

6.5.1.1 Challenges for validation of gridded products

Ozone_cci+ Level-3 limb profile data products are averages of single profile retrievals of a number limb and occultation instruments. Most products represent monthly mean data in 10° latitude bands, one product provides monthly means in smaller cells of 10° latitude by 20° longitude. In addition, some products contain profile data from multiple instruments. Single profile satellite data sets (so-called Level-2 data) are usually validated using space and time co-located reference data, but such an approach cannot simply be translated for aggregates of limb data (so-called Level-3 data). In the following sections, we describe an approach to evaluate the quality of these data records using ozonesonde as a reference.

The main challenge in evaluating Level-3 satellite data is the launch frequency and the spatial density of the ground-based network which introduces considerable (depending on the bin size of the satellite product) spatial and temporal sampling errors. Most stations launch one balloon per week or twice a month. There are only a handful of sites that perform more frequent soundings, all of which are located in Europe. It is therefore not expected that the small monthly sample of sonde observations is representative of the monthly mean state of the ozone field around the station, especially in winter months which exhibit larger geophysical variability. In addition, there are many latitude bands and latitude-longitude cells without or with just a few stations. Similarly, it is therefore not expected that the data from a handful of stations is representative of the mean state of 10°x20° grid cells and especially not for 10° zonal bands.

Nonetheless, the stratospheric ozone field correlates over several thousand km over several days [RD62]. A few 10°x20° grid cells in Europe and North America contain more than two stations with weekly soundings, which makes these prime locations to evaluate data quality. The investigation of larger-scale spatial structure of quality indicators in the stratosphere is more ambitious, especially for the zonally averaged products. The variability in the troposphere is larger than in the stratosphere, resulting in shorter correlation lengths and timescales [RD62]. Combined with increased measurement noise by limb sounders it is particularly challenging to assess satellite data quality in the lower part of the atmosphere.

6.5.1.2 Data preparation

Ozone_cci+ Level-3 limb profile products are compared to ozonesonde data on the same spatio-temporal grid. To this end, profile data of each ozonesonde station are first screened, unit converted, vertically smoothed & gridded and averaged over every month. The smoothing is done with a 1 km wide rectangular window. Uncertainties in the derived monthly mean value are reduced by rejecting months and grid levels with <2 (tropics) or <3 (higher latitudes) profiles. In a next step, these station monthly mean (SMM) data are used to derive the seasonal cycle at each site over the reference period 2004-2011 (same period as for satellite data). Seasonal cycle entries are discarded for months and grid levels that contain <4 years (tropics) or <5 years (higher latitudes) of SMM data. This requirement ensures a more accurate determination of the observed seasonal cycle, but is only satisfied for a select number of sites. Then, the relative anomaly with respect to this seasonal cycle is calculated, hereby removing the seasonal cycle and setting the average absolute level to zero over the reference period (as for some satellite products). This also removes instrument-related multiplicative offsets (i.e., bias). This data set is named the station monthly mean



anomaly (SMMa) data set (percent values). In a last step, the sonde data are mapped on the horizontal grid of the Ozone_cci+ products, either 10° zonal averages or smaller 10°x20° grid cells. This is done by averaging the SMM or SMMa data for the sites located in the Ozone_cci+ cell, hereby obtaining the cell monthly mean (CMM) and cell monthly mean anomaly (CMMa) data. All data are weighted equally, which effectively gives more weight to regions (Europe, North America) with more stations. Site-dependent instrument biases will generate in the station-averaged CMM data set not only random uncertainty but also jumps, due to differences in time coverage. However, such errors are suppressed for the station-averaged CMMa data set.

Single-sensor Ozone_cci+ Level-3 limb profile products contain either monthly mean ozone mole concentration on an altitude grid or monthly mean volume mixing ratio data on a pressure grid. The multi-sensor, merged products on the other hand consist of deseasonalized relative anomalies on an altitude grid. The estimation of some quality indicators (see next section) requires deseasonalized anomaly data for all single-sensor Ozone_cci+ data as well. For some instruments (GOMOS, MIPAS, SCIAMACHY, OSIRIS, ACE-FTS, SAGE II, OMPS-LP) the deseasonalized data are taken from the Ozone_cci+ product. For the remaining instruments (HALOE, SABER, Aura MLS) such data are delivered to us directly by the data providers. Below, satellite Level-3 data in ozone concentration units or VMR are referred to as *LP*, deseasonalized anomaly satellite data are referred to as *LPa*.

6.5.1.3 Estimation of data quality indicators

Quality indicators are computed from one of two comparison time series, both representing percentage differences. The first considers relative differences of satellite and sonde ozone concentrations (or volume mixing ratios) : $\Delta(\text{lat}/\text{lon},z,t) = 100 \times (LP(\text{lat}/\text{lon},z,t) - CMM(\text{lat}/\text{lon},z,t)) / CMM(\text{lat}/\text{lon},z,t)$. The second considers absolute differences of satellite and sonde deseasonalized relative anomaly data : $\Delta_o(\text{lat}/\text{lon},z,t) = LPa(\text{lat}/\text{lon},z,t) - CMMa(\text{lat}/\text{lon},z,t)$.

From these difference time series, three statistical indicators are computed.

- The median of the Δ (for single-sensor satellite data) or Δ_o (for merged satellite data) difference time series, as a proxy of the bias in the satellite product, albeit with the caveat that also representativeness differences and systematic errors in the ground-based data will contribute to this median difference.
- The spread in the Δ_o difference time series, derived as half the 16-84% interpercentile (which corresponds to the standard deviation of a normal distribution).
- The slope of a linear regression to the Δ_o difference time series as a proxy of the drift in the satellite data set, again with the caveat that representativeness differences and drifts in the ground-based data also contribute to this slope. We report the weighted average of the drift results over all cells. Furthermore, a χ^2 -approach is used to the scale drift uncertainties in order to incorporate unknown inhomogeneities in the sonde network [RD43].

Bias and comparison spread results will be shown below versus vertical coordinate (altitude or pressure) and latitude, drift results are only shown as a function of vertical coordinate.

6.5.2 Validation results

6.5.2.1 Bias

The bias of single-sensor Ozone_cci+ products is estimated as the median value of time series of the relative difference between gridded satellite and ozonesonde data. None of these records are deseasonalised prior to the comparison. As explained in Section 6.5.1.2, inhomogeneities (in space and time) present in the non-deseasonalised ozonesonde data records (CMM) lead to artificial variance of the satellite bias field in the horizontal domain. The bin-to-bin differences in the bias results visible in Figure 6.74 to Figure 6.76 (left



column) illustrate this issue. In most cases, such differences remain within the estimated 5-7% systematic uncertainty of ozonesonde data [RD79] but a few cells do stand out: e.g., 60°N-70°N for 25-30 km (Sodankylä), 50°N-60°N for 13-18 km (9 sites), 0°N-10°N for 20-30 km (Paramaribo, Sepang Airport) , 10°S-20°S for 10-16 km (Samoa, Suva) and 50°S-60°S for 22-30 km (Macquarie Island). The differences are similar for many satellite instruments which suggests that these are due to the ozonesonde data. In these cells, the bias results should not be blindly attributed to the satellite. Such issues are mostly avoided by deseasonalising the ozonesonde records prior to averaging in the horizontal domain (i.e. CMMa), although doing so also precludes the possibility to assess the bias of non-deseasonalised satellite data. Merged Ozone_cci+ products are reported as deseasonalised anomaly data, so their bias is calculated as the absolute difference between satellite and ground-based deseasonalised anomaly data. Since merged satellite and ozonesonde data are both deseasonalised using the same reference period, the absolute level of both time series should be, by construction, identical during 2004-2011. Deviations from zero bias can result from differences in sampling or from differences outside the reference period. Panels on the left of Figure 6.77 clearly show that the bin-to-bin variance in the horizontal domain is much reduced for median Δ_a , illustrating that station-to-station differences do not impact the bias analysis of anomaly data.

Most Level-3 bias results in Figure 6.74 to Figure 6.77 (left column) are in line with those obtained for the Level-2 validation analyses presented in Section 6.4. In the lower stratosphere, most gridded satellite and ozonesonde data differ on average by less than ~5%. SABER, MIPAS and ACE-FTS are clearly biased positive and exhibit larger biases between 5-10%. The updated MIPAS (v8) and ACE-FTS (v4) algorithm generally increasing the O3 concentrations by up to 5% which makes the positive bias larger compared to that of the previous data versions. The SCIAMACHY data are mostly too low and seem to exhibit a consistent oscillatory pattern in the vertical domain as well (e.g., less negative biases occur at 22 and 25 km). Vertical oscillations are also noted for Aura MLS in the extratropics at altitudes below the 80 hPa level, while in the tropics this feature may be masked by the larger percentage differences in the upper troposphere. Around the tropopause and below, the bias estimates are generally larger than 30% as a result of the lower sensitivity of limb and occultation measurements, but also due to the smaller ozone concentrations. Many satellite data records are biased high in the extratropical UTLS. In the tropics, the sign of the UTLS bias varies. HALOE data at altitudes below the 100-150 hPa level are clearly biased negative. There are no clear signs of hemispheric asymmetry in the stratospheric bias fields apart from the polar regions due to the Antarctic ozone hole season. South of 60°S the GOMOS, SCIAMACHY and ACE-FTS data may have more elevated positive biases of up to ~15% between 12-18 km. The Antarctic data by HALOE (and possibly SAGE II) appear more negative than in the Arctic. The latter results should not be overinterpreted as the low ozone concentrations in the Antarctic vortex and differences in spatio-temporal sampling of sonde and satellite instruments are expected to contribute considerably to the obtained bias estimates. Figure 6.77 shows very small biases for both merged satellite products, as anticipated. Biases remain less than 3-4% over the entire probed region of the atmosphere and there is no clear spatial structure apart from an 8% bias around the tropopause in the 10°-20°S band for the zonally averaged product.

6.5.2.2 Comparison spread

The structure of the spread (Figure 6.74 to Figure 6.77, middle column) in the comparison time series of Ozone_cci+ deseasonalized anomaly data is in line with that found for the Level-2 products (cfr. Section 6.4). Above 20 km, the observed spread is smallest in the tropics (~5%) and increases towards the polar regions (~12%). Below 20 km, the observed spread increases rapidly to 20-30% and more. Values in the Antarctic lowermost stratosphere are clearly larger than in the Arctic. As mentioned before, the variance in the comparisons receives contributions from natural variability due to differences in spatio-temporal sampling of the satellite and ground-based data. This leads to Level-3 spreads that are 1-2% larger than at Level-2 for all records except SCIAMACHY. For SCIAMACHY, much smaller comparison spreads are noted at Level-3 when compared to the Level-2 data, as a result of the temporal averaging at Level-3. Also, the impact of sampling



differences increases where variability in the atmospheric ozone field is larger, e.g., in the polar regions, in the UTLS, in the winter season at mid-latitudes, or during the Antarctic ozone hole season. Such spatial dependence of the dispersion is indeed noted in the comparisons, so in these parts of the atmosphere the observed spread cannot be blindly attributed to lower quality of the satellite products. However, from this analysis, it cannot be excluded that satellite random uncertainty degrades in these regions, e.g., due to lower signal to noise ratios. Comparison spreads in the analyses of the new ACE-FTS and MIPAS data records are equal to those found for earlier data versions. This implies that the algorithm update for these sensors does not change the random uncertainty.

6.5.2.3 Long-term stability

Also the drift estimates for the Level-3 satellite products (Figure 6.74 to Figure 6.77, right column) are in accordance with those of the underlying Level-2 data sets, though the uncertainty of Level-3 drift is often considerably smaller than at Level-2 (cfr. Section 6.4). The statistics available for the analysis of Level-2 data is more limited due to the stringent spatio-temporal co-location requirements. The Level-3 analysis considers the entire time series of both ground-based and satellite records which increases the ability to constrain the regression.

No significant drift relative to the sonde record is noted for most single-sensor satellite products. Exceptions are a negative drift of GOMOS data below ~25 km (up to 5% per decade) and in HALOE data above 24 km (~5% per decade). A significant positive drift is clearly noted in SCIAMACHY data (1-3% per decade). The stability of OMPS-LP data changes by 0.5-1% per decade w.r.t. the Feb 2020 data release. Current drift estimates for the OMPS-LP data record in the lower stratosphere are smaller than 2% per decade and become insignificant over (at least) the 22-30 km vertical range. The drift of SABER data is estimated at 1-2% per decade above the 50 hPa level. The SABER drift estimates extend into the UTLS and indicate a positive drift of 2-3% per decade. Above 25 km, almost all recent satellite products appear to drift by ~2% per decade to more positive values, a feature that was also seen in analyses of Level-2 data [RD43]. It is not excluded that this may be related to instabilities in the ground-based data, rather than an issue in the satellite data records. For instance, the dropoff identified by Stauffer et al. [RD83] at some ozonesonde sites contributes to additional positive satellite drift estimates. Below about 20 km, it is very difficult to assess the significance of the results as the uncertainties are most likely underestimated. Further research may lead to more solid estimates. Therefore, reports of a large drift values for HALOE, OSIRIS, SABER, SCIAMACHY and OMPS-LP below 15 km are indicative, but not conclusive.

The merged zonally averaged product (SAGE-CCI-OMPS) is stable between 15-30 km. Below 15 km, the negative 2% per decade drift is indicative, but most likely statistically insignificant for reasons mentioned before. The results in individual cells of the merged latitude-longitude resolved product (MEGRIDOP) are scattered due to residual inhomogeneities in the ground-based data records (Figure 6.77). As a result, drift estimates in individual cells are subject to large uncertainties which make it challenging to perform a robust and precise assessment of the stability of this particular Ozone_cci+ product. Estimates in most cells are less than 3% per decade between 20-30 km. MEGRIDOP drift is generally positive when averaged over the ground-network. The drift is significant between 15-30 km and lies between 0.5-2.5% per decade. However, it may be that significance is overestimated due to too low drift uncertainty estimates.

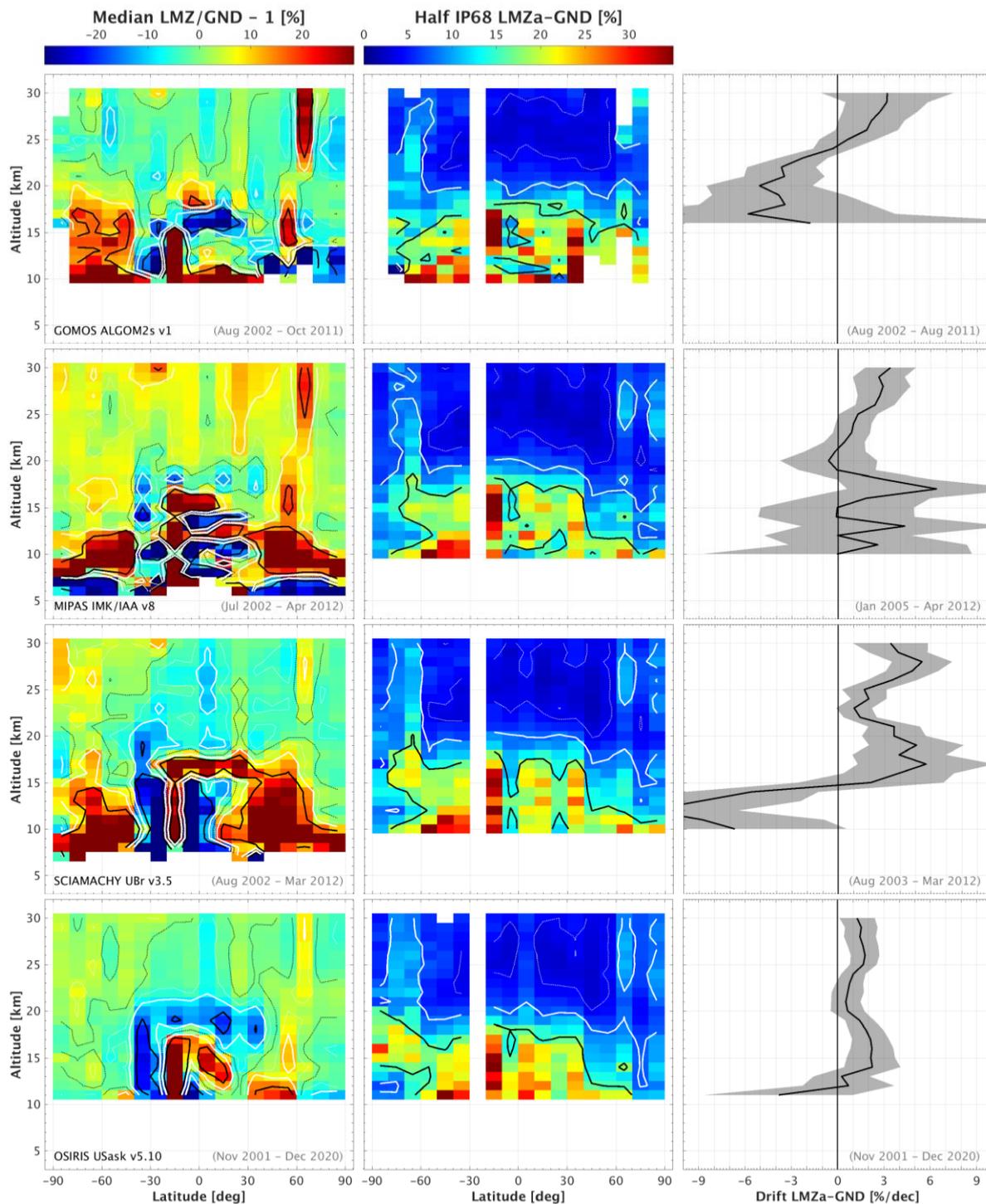


Figure 6.74 - Spatial structure of median (left), spread (centre) and drift (right, with 95% CI) in comparisons of four Ozone_cci+ Level-3 limb profile products to NDACC/GAW/SHADOZ ozonesonde data : GOMOS, MIPAS, SCIAMACHY and OSIRIS. The MIPAS anomaly data (LMZa) cover only the last part of the mission (2005-2012). Contour lines show the 0% (dotted black), 4% (dotted white), 8% (solid white) and 16% levels (solid black). Positive values indicate that satellite data are larger than the reference data.

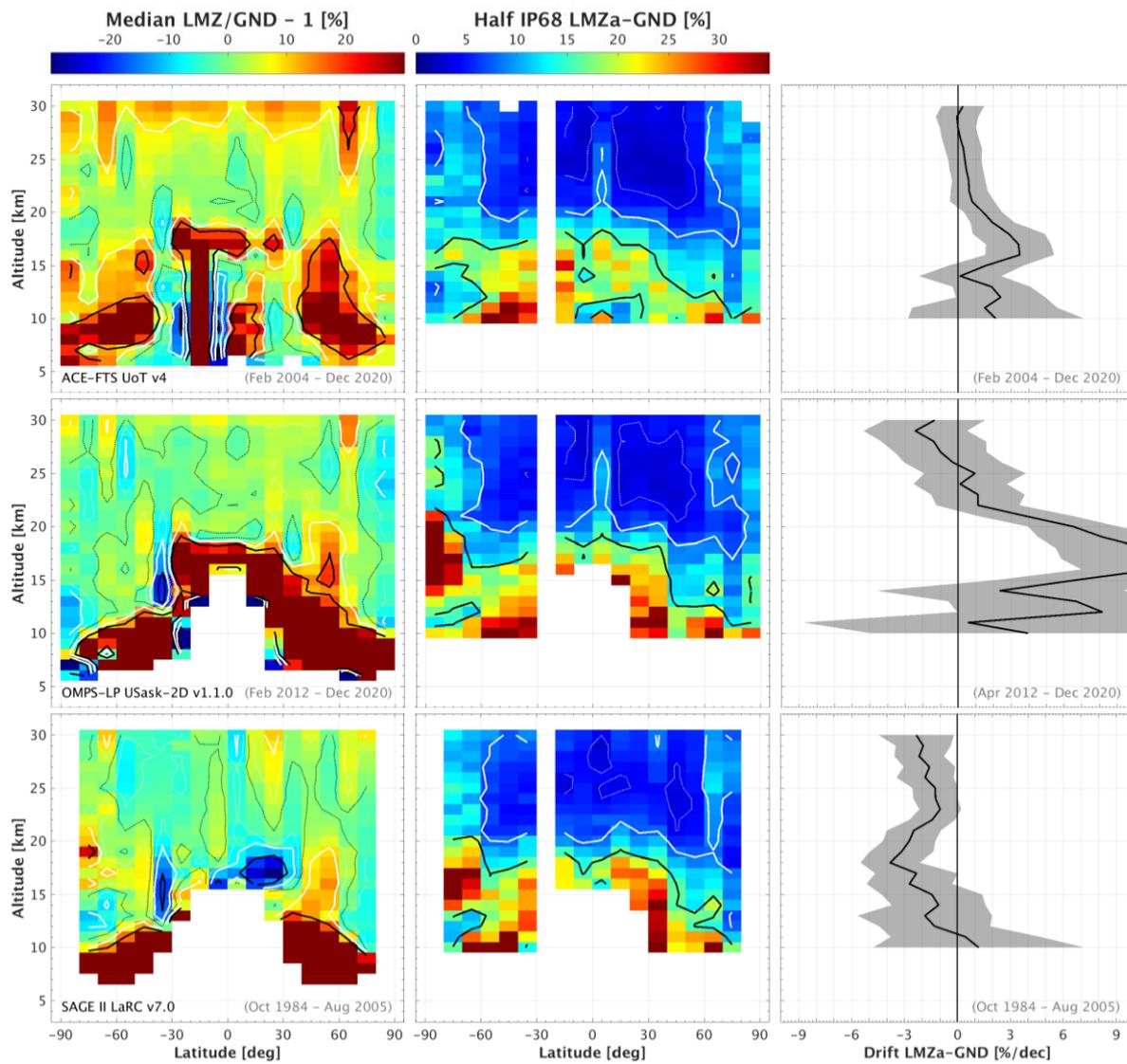


Figure 6.75 - As in Figure 6.74, for ACE-FTS, OMPS-LP and SAGE II Ozone_cci+ products (top to bottom).

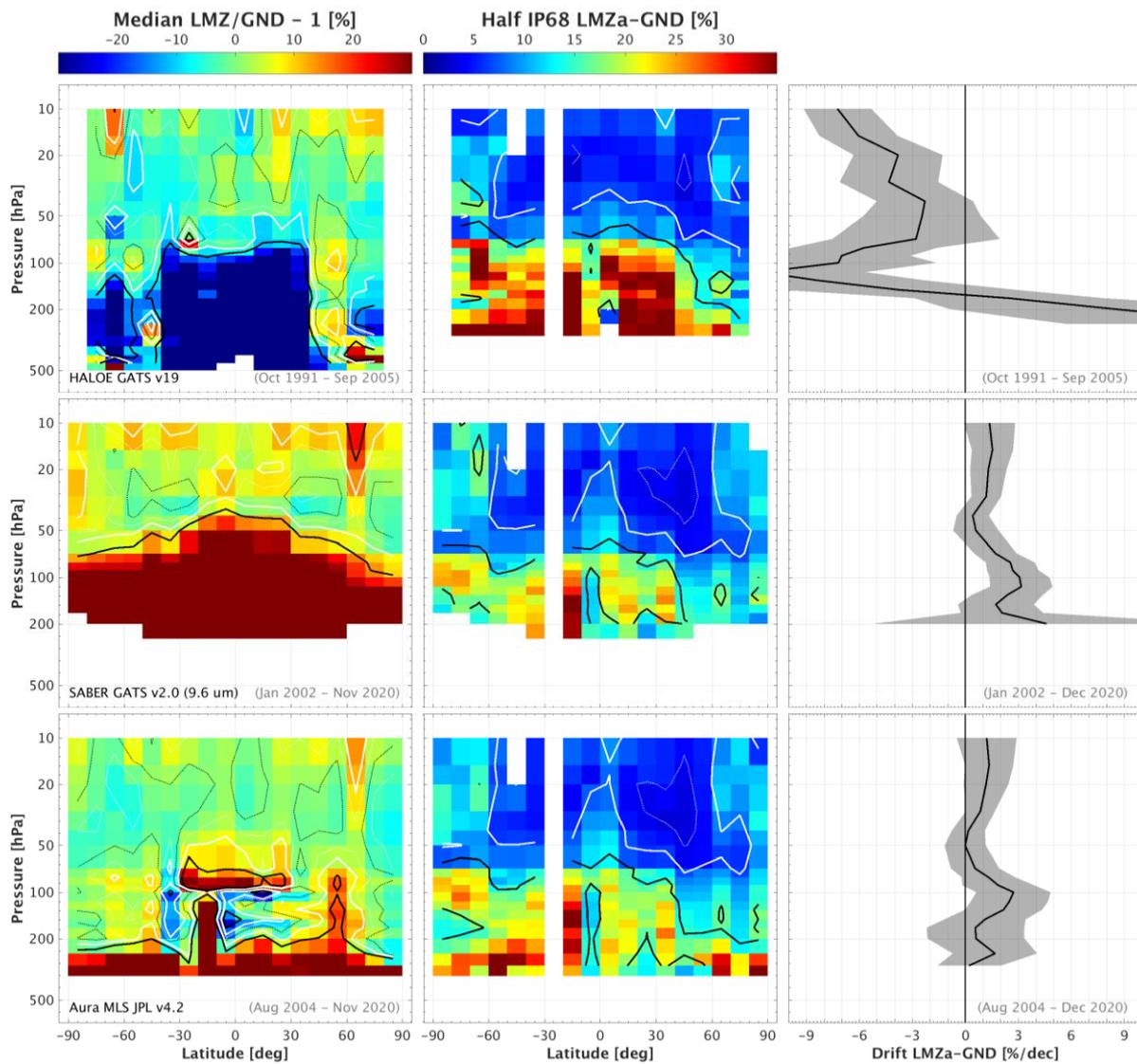


Figure 6.76 - As in Figure 6.74, for the HALOE, SABER and Aura MLS Ozone_cci+ products, reported as volume mixing ratios on a pressure grid (top to bottom).

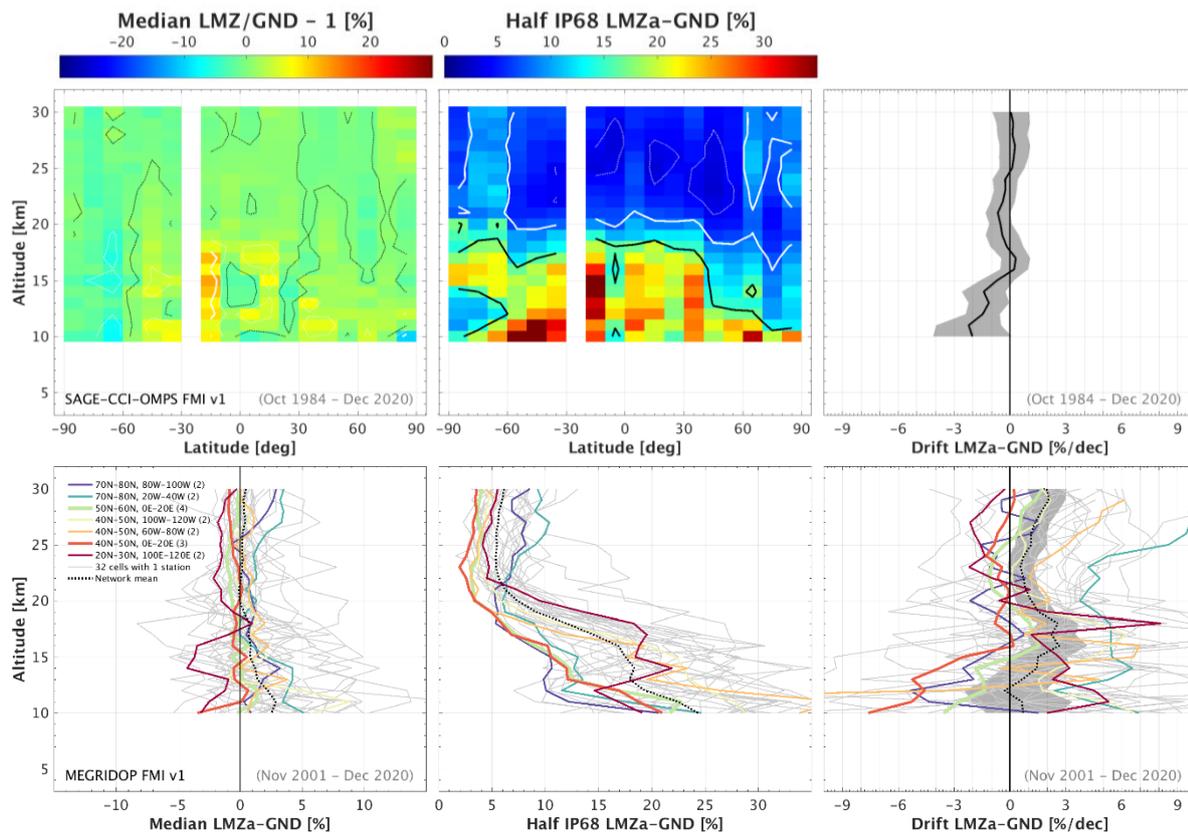


Figure 6.77 - Structure of median (left), spread (centre) and drift (right) in comparisons of the merged latitude resolved (top) and latitude-longitude resolved (bottom) Ozone_cci+ products to NDACC/GAW/SHADOZ ozonesonde data. Contour lines show the 0% (dotted black), 4% (dotted white), 8% (solid white) and 16% levels (solid black). Positive values indicate that CCI data are larger than the reference data. Coloured lines in the bottom panels indicate grid cells with at least two stations, those with one site are shown in grey. The black dotted curve shows the mean value of the quality indicator over the network (the shaded area is 95% CI). The time range of the analysis of deseasonalised anomaly ozone data is mentioned in grey.



6.5.3 Compliance with user requirements

A summarizing overview of the compliance of the Ozone_cci+ Level-3 limb ozone profile products with user requirements (URD v3.1 [RD8]) is presented in Table 6.36. Caution: comparison spread in polar regions and the UTLS is likely dominated by sampling differences. It can therefore not be excluded that, in these regions, the satellite products are compliant with the requirements.

Table 6.36 - Compliance of Ozone_cci+ Level-3 limb profile data products with user requirements (URD v3.1). Green = fully compliant; orange = one of more products may be not compliant; red = no products compliant.

Quantity	Requirement		Compliance assessment	
	Lower Stratosphere (LS)	Middle Atmosphere (MA)	Occultation	Limb
Horizontal resolution	100 – 200 km	200 – 400 km		Uncertain, but expected compliant in MA, not in LS
Vertical resolution	1 – 2 km	2 – 4 km		Compliant in MA, not in LS
Observation frequency	Daily – weekly	Daily – weekly	Only 30 solar occultation profiles per day, but more for stellar occultation	
Time period	(1980-2010) – (2003-2010)	(1980-2010) – (2003-2010)		
Accuracy in height attribution	±500 m	±500 m		Uncertain, expected close to compliant

Altitude range	...-20 km			20-30 km			20-30 km
Requirement	Random uncertainty < 8-16%			Random uncertainty < 8%			Drift < 1-3% / dec.
Latitude	Polar	Mid-lat.	Tropics	Polar	Mid-lat.	Tropics	Global
LP_L3_GOMOS							
LP_L3_MIPAS							
LP_L3_SCIA							
LP_L3_OSIRIS							
LP_L3_ACE							
LP_L3_OMPS							
LP_L3_SAGE-II							
LP_L3_HALOE							
LP_L3_SABER							
LP_L3_MLS							
LP_L3_MERGED_SAGE_CCI_OMPS							
LP_L3_MEGRIDOP							



7 Comparison error budget and compliance criteria

In the validation work reported in the previous sections, differences between satellite and reference measurements (and the statistics on those differences, such as means, medians, and spreads) are compared to user requirements to verify whether the data are fit-for-purpose. However, from a metrological (i.e., measurement science) point-of-view, a crucial test is whether a measurement is accurate to the level indicated by its reported uncertainty. While the reporting of uncertainties is becoming more and more common place in the ozone community, these uncertainties are only rarely used in the validation work. This was recognized in the requirements listed for the current Ozone_cci+ project, more specifically in technical requirement 23 (TR-23), which states: “Particular attention will be paid to an evaluation of the ex-ante uncertainties reported with the ECV data products.” The current section therefore deals with an assessment of the ozone measurements, not w.r.t. the user requirements, but with respect to their reported (ex-ante) uncertainties.

Below, we first look at a simple formalism to verify the consistency of two measurements with respect to their reported measurement uncertainties, as used by the GRUAN (GCOS Reference Upper Air Network) community in the validation of satellite temperature and humidity measurements with radiosonde soundings. In the subsequent section, we show that this consistency test is in fact a much-simplified version of the classical χ^2 (“Chi Square”) test on a set of differences, as typically used in various modelling communities to test the agreement between model and measurement. This test is in principle straightforward to apply to the validation of satellite ozone measurements if both satellite and reference measurements come with ex-ante uncertainty estimates. As a first application, we perform the χ^2 test on comparisons of S5p-TROPOMI total ozone with ground-based measurements. We end this section with some caveats and prospects on further applications and improvements.

7.1 Consistency of two measurements

Assuming we have two measurements, m_1 and m_2 , with respective reported uncertainties u_1 and u_2 , the baseline consistency is to compare the difference between both measurements with the quadratic sum of the uncertainties:

$$|m_1 - m_2| < k \sqrt{u_1^2 + u_2^2}. \quad (7.1)$$

Depending on the coverage factor k that is needed to make $|m_1 - m_2|$ smaller than $k \sqrt{(u_1^2 + u_2^2)}$, one can decide on the level of consistency, following Table 7.1. When the co-location is not perfect, due to differences in spatiotemporal smoothing and sampling, one can add an additional component, the co-location mismatch uncertainty to the quadratic sum (if quantifiable in terms of a variance). More details on that are provided in Section 7.4.

Table 7.1 - Consistency assessment for two measurements, m_1 and m_2 , with reported uncertainties u_1 and u_2 , from Immler et al. (2010) [RD44].

$ m_1 - m_2 < k \sqrt{u_1^2 + u_2^2}$	TRUE	FALSE	significance level
k=1	consistent	suspicious	32%
k=2	in agreement	significantly different	4.5%
k=3	-	inconsistent	0.27%



7.2 Consistency of a set of N co-located measurement pairs

When analysing a set of co-located measurement pairs, it is interesting to have a quantifier for the consistency for the entire sample. As described in Von Clarmann (2006), which builds on Rodgers (2003) but is actually an application of the work by Pearson (1900), the χ^2 test is a natural extension of the 2-measurement consistency test. It also allows for a separation between systematic (constant for the whole sample) and random components. To that end, the mean difference (bias, \bar{b}_{diff}) should be calculated first and (1) compared against the combined systematic uncertainty in both data sets, and (2) subtracted from the individual differences before proceeding to calculate a χ^2 . The latter should then use only the random uncertainties reported with the measurements. The generic formalism as presented by Von Clarmann (2006) looks as follows:

$$\langle \chi^2 \rangle = \langle (\hat{x}_{val} - \hat{x}_{ref} - \bar{b}_{diff})^T S_{diff,random}^{-1} (\hat{x}_{val} - \hat{x}_{ref} - \bar{b}_{diff}) \rangle \quad (7.2)$$

with an expected value of $N-1$ (N the number of measurement pairs). $S_{diff,random}^{-1}$ represents the combined random uncertainty (as a covariance matrix). The application to S5p-TROPOMI data in the next section will clarify how to this is implemented in practice.

7.3 Application to S5p-TROPOMI total O₃

As a first application, we compare here S5p-TROPOMI OFFL total ozone columns with ground-based reference measurements (Dobson, Brewer, and SAOZ) and verify their consistency in terms of reported uncertainties.

7.3.1 A case study: S5p vs. the Dobson at Brisbane, AUS

In Figure 7.1, we present the time series of co-located S5p and Dobson total ozone column measurements at Brisbane, Australia, including their reported uncertainties.

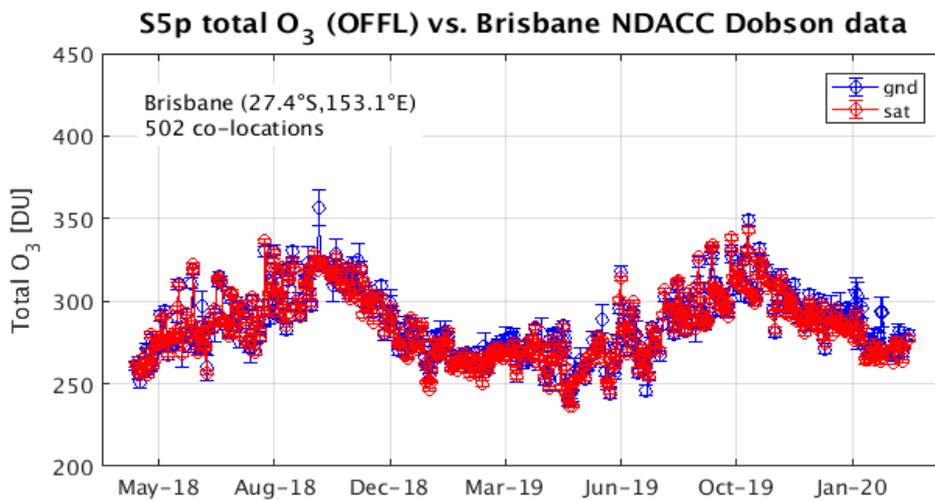


Figure 7.1 - Time series of S5p and Dobson total ozone measurements, with their reported uncertainties, at Brisbane, Australia.

From these time series, we can calculate the differences, bias-correct them using the mean difference, and “normalize” them by the quadratically combined measurement uncertainties. Summing these over the entire time series yields the χ^2 :

$$\chi^2 = \sum \frac{(SAT - GND - mean(SAT - GND))^2}{\sigma_{SAT}^2 + \sigma_{GND}^2} \quad (7.3)$$



As the expected value of this Chi Square depends on the sample size (N-1), it can aid interpretation to calculate a Reduced Chi Square, χ_r^2 :

$$\chi_r^2 = \frac{1}{N-1} \sum \frac{(SAT-GND - mean(SAT-GND))^2}{\sigma_{SAT}^2 + \sigma_{GND}^2} \quad (7.4)$$

which has an expectation value of 1. This is further illustrated for this particular case study in Figure 7.2. The χ_r^2 of 1.3 indicates excellent consistency within the reported uncertainties. This is further confirmed by the number of pairs satisfying the different coverage factors of the pair-wise test, which is close to what would be expected for a Normal distribution of the errors.

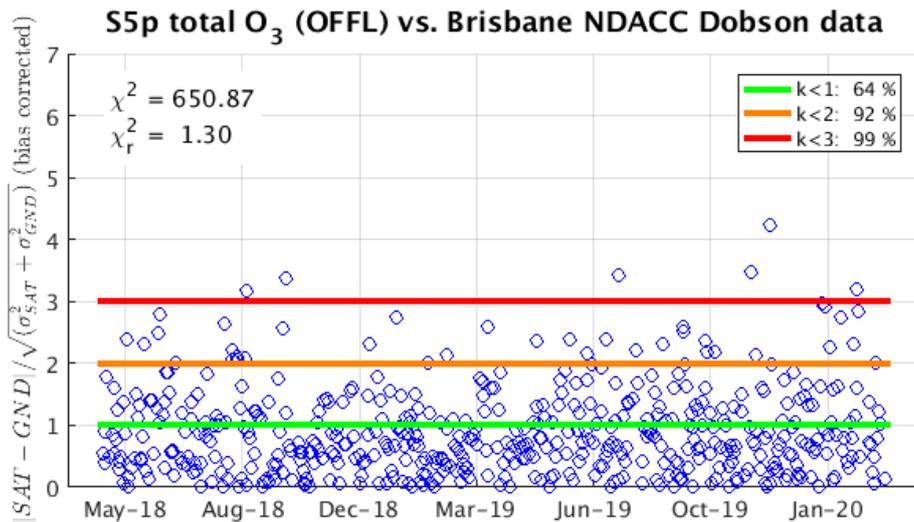


Figure 7.2 - Blue markers: Absolute differences between co-located S5p and Dobson total ozone columns at Brisbane, bias-corrected and normalized by their combined uncertainty. The sum of these constitutes the $C\chi^2$, which, divided by the number of co-locations (502 here) minus 1, yields the χ_r^2 . The coloured solid lines represent the criteria from the 2-measurement consistency test described in Section 7.1 and Table 7.1.

7.3.2 Network-wide results

In the previous section, we derived the χ_r^2 for the time series of differences at a single station. Applying this for the entire network of ground-based reference instruments, i.e. deriving a χ_r^2 for every site in the network, allows us to assess the measurements and their uncertainties under various conditions. The global distribution of the instruments used here is visualized in Figure 7.3.

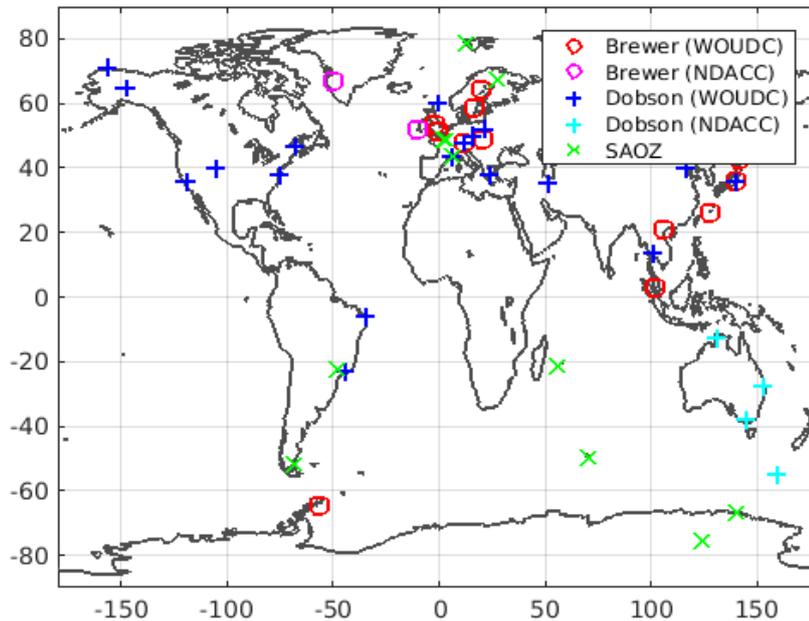


Figure 7.3 - Global distribution of the ground-based reference instruments used for the assessment of S5p OFFL total ozone columns and their reported uncertainties.

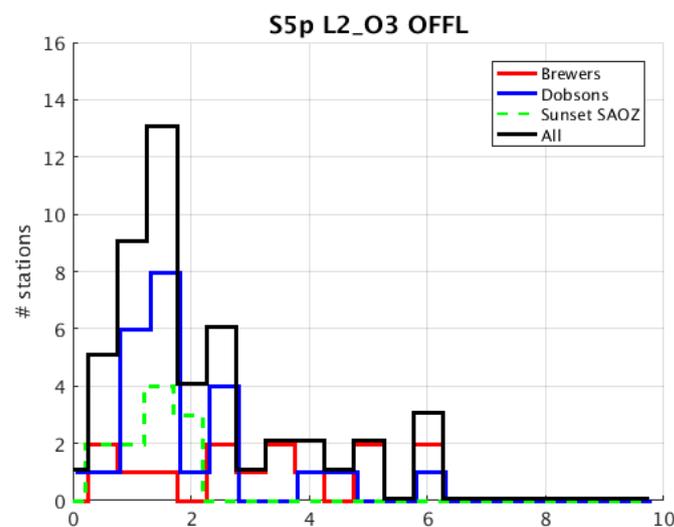


Figure 7.4 - Histogram of the χ_r^2 values (one per reference instrument) for S5p OFFL total ozone versus different ground-based instruments.

A histogram of the χ_r^2 values each instrument in the network is presented in Figure 7.4. The theoretical expectation value of the χ_r^2 is one. The mode of the distribution is close to one, indicating overall good consistency between satellite and ground-based measurements in the sense that the differences are in line with their reported uncertainties. The width of the expected distribution depends on the number of degrees of freedom (i.e. on the number of co-locations), so it is not straightforward to quantify this for this sample of χ_r^2 values where each has an underlying distribution with a different number of degrees of freedom. Expert judgement suggest values up to 3 to be acceptable. Larger values (unless based on very few co-locations)



deserve further investigation. Some of the outliers in Figure 7.4 could already be traced back to issues in the ground-based data. This is work in progress.

7.4 Caveats and prospects

The analysis presented above should be seen mostly as a proof-of-concept, as a few aspects need further attention:

- The separation between random and systematic uncertainty components is not always feasible as no such distinction is made in the reported uncertainties. For some (ground-based) instruments, only a total uncertainty is provided, others provide separate uncertainties from different error sources, but it is up to the user to judge to what extent these act systematic or random on the (time) scale of the analysis.
- The uncertainty on the ground-based data is not always small enough to test truly the S5p uncertainties. For example, the S5p OFFL total ozone column uncertainties are typically 4-5 times smaller than those of the SAOZ measurements are, and they therefore hardly feature in the error budget (where uncertainties are combined quadratically).
- The additional uncertainty due to co-location mismatch is not yet taken account in the analysis. It can be done e.g. with model-based OSSEs (see Verhoelst et al., 2015) or – somewhat simplified – using an additional variance from e.g. the co-location mismatch uncertainty tables produced in the H2020 GAIA-CLIM project (Verhoelst et al., 2017). Yet apparently, it plays only a minor role for the current case study, in which we already went to great lengths to minimize co-location mismatch (e.g. by using an observation operator based co-location scheme for the twilight zenith-sky ground-based measurements).

Ideally, these concerns are addressed before extending the analysis to L2 data sets from other sensors (both columns and profiles). The principle can also be adapted for use on L3 and L4 data sets, where it would – amongst other possibilities – provide an elegant solution to check the consistency between punctual reference measurements and spatiotemporally averaged satellite measurements, provided the satellite data sets contain information on measurement (or model) variance within a grid cell.



8 References

8.1 *Applicable documents*

- [RD1] CCI+ SOW: ESA Climate Change Initiative Extension (CCI+) Phase 1: New R&D on CCI ECVs – Statement of Work, Ref. ESA-CCI-EOPS-PRGM-SOW-18-0118, Issue 1 Revision 6, 31/05/2018.
- [RD2] CCI+ Ozone Technical Proposal: ESA Ozone_cci+ Technical Proposal – Essential Climate Variable (ECV) - Ozone, Proposal to ESA in response to ITT AO/1-9322/18/I-NB, 14/09/2018.
- [RD3] CCI+ Ozone Product Validation Plan: ESA Ozone_cci+ Product Validation Plan (PVP), Ref. Ozone_cci+_PVP_2.1, Issue 2, Revision 1, 06/12/2020.
- [RD4] CCI+ Ozone Algorithm Theoretical Basis Document: ESA Ozone_cci+ Algorithm Theoretical Basis Document (ATBD), Ref. Ozone_cci+_D2p1_ATBD_2, Issue 2, Revision 0, 05/11/2021.

8.2 *Reference documents*

8.2.1 **User requirements**

- [RD5] CMUG: User Requirement Document | Meeting the needs of the Climate Community - Requirements, Deliverable 1.1, Climate Modelling User Group, Issue 2, Revision 1, 11/01/2021.
- [RD6] DARD: Ozone CCI Phase II Data Access Requirement Document (DARD), Issue 2, Revision 1, Ref. Ozone_cci_DARD_1.1, 25/05/2016.
- [RD7] IGACO: The changing atmosphere. An integrated global atmospheric chemistry observation theme for the IGOS partnership. Report of the Integrated Global Atmospheric Chemistry Observation (IGACO) theme team, Ref. ESA SP-1282, GAW No. 159, WMO-TD No. 1235, 2004.
- [RD8] URD: Ozone CCI User Requirement Document, Ref. Ozone_cci_URD_3.1, Issue 3, Revision 1, 05/03/2021.
- [RD9] WMO Observing Systems Capabilities Analysis and Review tool (OSCAR) available on-line from <http://www.wmo-sat.info/oscar/observingrequirements>

8.2.2 **International standards and frameworks**

- [RD10] CDRH: Center for Devices and Radiological Health (CDRH), General Principles of Software Validation; Final Guidance for Industry and FDA Staff, CBER CDRH/OC Doc. N. 938, January 11, 2002. Publicly available via <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance>
- [RD11] CEOS: Committee on Earth Observation Satellites (CEOS): Terms and Definitions and other documents and resources, publicly available on <http://calvalportal.ceos.org>
- [RD12] GUM: Joint Committee for Guides in Metrology (JCGM/WG 1) 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in a measurement (GUM), http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
- [RD13] Larssen, S., R. Sluyter, and C. Helms, Criteria for EUROAIRNET – The EEA Air Quality Monitoring and Information Network, 1999.
- [RD14] Nappo, C.J., Caneill J.Y., Furman R.W., Gifford F.A., Kaimal J.C., Kramer M.L., Lockhart T.J., Pendergast M.M, Pielke R.A., Randerson D., Shreffler J.H., and Wyngaard J.C., The Workshop on the Representativeness of Meteorological Observations, June 1981, Boulder, CO, Bull. Am. Meteorol. Soc. 63, 761-764, 1982.



- [RD15] NIST: Prokhorov, A. V., R. U. Datla, V. P. Zakharenkov, V. Privalsky, T. W. Humpherys, and V. I. Saprisky, Spaceborne Optoelectronic Sensors and their Radiometric Calibration. Terms and Definitions. Part 1. Calibration Techniques, Ed. by A. C. Parr and L. K. Issaev, NIST Technical Note NISTIR 7203, March 2005.
- [RD16] VIM: Joint Committee for Guides in Metrology (JCGM/WG 2) 200:2008 & ISO/IEC Guide 99-12:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), <http://www.bipm.org/en/publications/guides/vim.html>
- [RD17] WMO Quality Management Framework (QMF), home page at <http://www.wmo.int/pages/prog/www/QMF-Web/home.html>
- [RD18] QA4EO – A Quality Assurance framework for Earth Observation, established by the CEOS. It consists of ten distinct key guidelines linked through an overarching document (the http://qa4eo.org/docs/Guidelines_Framework_v3.0.pdf QA4EO Guidelines Framework) and more community-specific QA4EO procedures, all available on <http://qa4eo.org/documentation.html> A short QA4EO "user" guide has been produced to provide background into QA4EO and how one would start implementing it (http://qa4eo.org/docs/QA4EO_guide.pdf)
- [RD19] ISO Quality Management Principles available at <http://www.iso.org/iso/iso9000-14000/understand/qmp.html>
- [RD20] NetCDF Climate and Forecast Metadata Convention, <http://cf-psmdi.llnl.gov>
- [RD21] Fahre Vik, A., T. Krognæs, S-E. Walker, S. Bjørndalsæter, C. Stoll, T. Bårde, R. Paltiel, and B. Glosli, ESA Campaign Database (CDB) user manual, NILU Technical Note O-103045, 100 pp., April 2006. http://nadir.nilu.no/cdb/doc/CDB_manual_20060405.pdf
- [RD22] World Meteorological Organization, WMO Global Atmosphere Watch (GAW) Strategic Plan: 2008-2015, GAW Report No. 172 / WMO TD No. 1384.

8.2.3 Validation measurements, validation methods and assessments

- [RD23] Adams, C., A. E. Bourassa, A. F. Bathgate, C. A. McLinden, N. D. Lloyd, C. Z. Roth, E. J. Llewellyn, J. M. Zawodny, D. E. Flittner, G. L. Manney, W. H. Daffer, and D. A. Degenstein. Characterization of Odin-OSIRIS ozone profiles with the SAGE II dataset. *Atmos. Meas. Tech.*, 6 (5): 1447–1459, doi: 10.5194/amt-6-1447-2013, 2013.
- [RD24] Adams, C., A. E. Bourassa, V. Sofieva, L. Froidevaux, C. A. McLinden, D. Hubert, J.-C. Lambert, C. E. Sioris, and D. A. Degenstein. Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozonesondes. *Atmos. Meas. Tech.*, 7 (1): 49–64, doi: 10.5194/amt-7-49-2014, 2014.
- [RD25] Backus, G.E., and F. Gilbert, Uniqueness in the Inversion of inaccurate Gross Earth Data, *Philosophical Transactions of the Royal Society of London A*, 266, 123-192, 1970.
- [RD26] Balis, D., J.-C. Lambert, M. Van Roozendaal, D. Loyola, R. Spurr, Y. Livschitz, P. Valks, V. Amiridis, P. Gerard, and J. Granville, Ten years of GOME/ERS-2 total ozone data – The new GOME Data Processor (GDP) Version 4: II Ground-based validation and comparisons with TOMS V7/V8, *Journal of Geophysical Research – Atmosphere*, Vol. 112, D07307, doi:10.1029/2005JD006376, 2007.
- [RD27] Bernhard, G., R. D. Evans, G. J. Labow, and S. J. Oltmans, Bias in Dobson total ozone measurements at high latitudes due to approximations in calculations of ozone absorption coefficients and air mass, *J. Geophys. Res.*, 110, D10305, doi:10.1029/2004JD005559, 2005.



- [RD28] Bourassa, A. E., Roth, C. Z., Zawada, D. J., Rieger, L. A., McLinden, C. A., and Degenstein, D. A.: Drift-corrected Odin-OSIRIS ozone product: algorithm and updated stratospheric ozone trends, *Atmos. Meas. Tech.*, 11, 489–498, <https://doi.org/10.5194/amt-11-489-2018>, 2018.
- [RD29] Boynard, A., Hurtmans, D., Koukouli, M. E., Goutail, F., Bureau, J., Safieddine, S., Lerot, C., Hadji-Lazaro, J., Wespes, C., Pommereau, J.-P., Pazmino, A., Zyrichidou, I., Balis, D., Barbe, A., Mikhailenko, S. N., Loyola, D., Valks, P., Van Roozendaal, M., Coheur, P.-F., and Clerbaux, C.: Seven years of IASI ozone retrievals from FORLI: validation with independent total column and vertical profile measurements, *Atmos. Meas. Tech.*, 9, 4327–4353, doi:10.5194/amt-9-4327-2016, 2016.
- [RD30] Boynard, A., D. Hurtmans, K. Garane, F. Goutail, J. Hadji-Lazaro, M. E. Koukouli, C. Wespes, C. Vigouroux, A. Keppens, J.-P. Pommereau, A. Pazmino, D. Balis, D. Loyola, P. Valks, R. Sussmann, D. Smale, P.F. Coheur, and C. Clerbaux: Validation of the IASI FORLI/EUMETSAT O3 products using satellite (GOME-2), ground-based (Brewer-Dobson, SAOZ, FTIR) and ozonesonde measurements, *Atmos. Meas. Tech.*, <https://doi.org/10.5194/amt-11-5125-2018>, 2018.
- [RD31] Bramstedt, K., Stone, T. C., Gottwald, M., Noël, S., Bovensmann, H., and Burrows, J. P.: Improved pointing information for SCIAMACHY from in-flight measurements of the viewing directions towards sun and moon, *Atmos. Meas. Tech.*, 10, 2413–2423, doi:10.5194/amt-10-2413-2017, 2017.
- [RD32] Burkholder, J.B., and R.K. Talukdar, Temperature dependence of the ozone absorption spectrum over the wavelength range 410 to 760 nm, *Geophysical Research Letters*, Volume 21, Issue 7, pages 581–584, doi:10.1029/93GL02311, 1994.
- [RD33] Calisesi, Y., V. T. Soebijanta, and R. van Oss, Regridding of remote soundings: Formulation and application to ozone profile comparison, *J. Geophys. Res.*, 110, D23306, doi:10.1029/2005JD006122, 2005.
- [RD34] Coldewey-Egbers, M., D.G. Loyola, M. Koukouli, D. Balis, J.-C. Lambert, T. Verhoelst, J. Granville, M. van Roozendaal, C. Lerot, R. Spurr, S.M. Frith, and C. Zehner, The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative, *Atmos. Meas. Tech.*, 8, 3923–3940, 2015.
- [RD35] Dufour, G., Eremenko, M., Griesfeller, A., Barret, B., LeFlochmoën, E., Clerbaux, C., Hadji-Lazaro, J., Coheur, P.-F., and Hurtmans, D.: Validation of three different scientific ozone products retrieved from IASI spectra using ozonesondes, *Atmos. Meas. Tech.*, 5, 611–630, doi:10.5194/amt-5-611-2012, 2012.
- [RD36] Eckert, E., von Clarmann, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Leblanc, T., McDermid, S., Pastel, M., Steinbrecht, W., Swart, D. P. J., Walker, K. A., and Bernath, P. F.: Drift-corrected trends and periodic variations in MIPAS IMK/IAA ozone measurements, *Atmos. Chem. Phys.*, 14, 2571–2589, doi:10.5194/acp-14-2571-2014, 2014.
- [RD37] Frith, S. M., Stolarski, R. S., Kramarova, N. A., and McPeters, R. D.: Estimating uncertainties in the SBUV Version 8.6 merged profile ozone data set, *Atmos. Chem. Phys.*, 17, 14695–14707, doi:10.5194/acp-17-14695-2017, 2017.
- [RD38] Garane, K., Lerot, C., Coldewey-Egbers, M., Verhoelst, T., Koukouli, M. E., Zyrichidou, I., Balis, D. S., Danckaert, T., Goutail, F., Granville, J., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., Pommereau, J.-P., Van Roozendaal, M., and Zehner, C.: Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 1: Ground-based validation of total ozone column data products, *Atmos. Meas. Tech.*, 11, 1385–1402, <https://doi.org/10.5194/amt-11-1385-2018>, 2018.



- [RD39] Garane, K., Koukouli, M.-E., Verhoelst, T., Lerot, C., Heue, K.-P., Fioletov, V., Balis, D., Bais, A., Bazureau, A., Dehn, A., Goutail, F., Granville, J., Griffin, D., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., McLinden, C., Pazmino, A., Pommereau, J.-P., Redondas, A., Romahn, F., Valks, P., Van Roozendael, M., Xu, J., Zehner, C., Zerefos, C., and Zimmer, W.: TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions, *Atmos. Meas. Tech.*, 12, 5263–5287, <https://doi.org/10.5194/amt-12-5263-2019>, 2019.
- [RD40] Godin, S., A.I. Carswell, D.P. Donovan, H. Claude, W. Steinbrecht, I.S. McDermid, T.J. McGee, M.R. Gross, H. Nakane, D.P.J. Swart, H.B. Bergwerff, O. Uchino, P. von der Gathen, and R. Neuber, Ozone differential absorption lidar algorithm intercomparison, *Appl. Opt.* 38, 6225-6236, 1999.
- [RD41] Godin, S., Bergeret, V., Bekki, S., David, C., and Mégie, G.: Study of the interannual ozone loss and the permeability of the Antarctic polar vortex from aerosol and ozone lidar measurements in Dumont d’Urville (66.4S, 140E), *J. Geophys. Res.*, 106, 1311-1330, doi:10.1029/2000JD900459, 2001.
- [RD42] Hendrick, F., Pommereau, J.-P., Goutail, F., Evans, R. D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M., and Van Roozendael, M.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations, *Atmos. Chem. Phys.*, 11, 5975-5995, doi:10.5194/acp-11-5975-2011, 2011.
- [RD43] Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., van Gijssel, J. A. E., Van Malderen, R., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, *Atmos. Meas. Tech.*, 9, 2497-2534, doi:10.5194/amt-9-2497-2016, 2016.
- [RD44] Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, *Atmos. Meas. Tech.*, 3, 1217–1231, <https://doi.org/10.5194/amt-3-1217-2010>, 2010.
- [RD45] Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., et al., Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change, *J. Environ. Monit.*, 6, 721-733. doi:10.1039/B404256E, 2004.
- [RD46] Keppens, A., Lambert, J.-C., Granville, J., Miles, G., Siddans, R., van Peet, J. C. A., van der A, R. J., Hubert, D., Verhoelst, T., Delcloo, A., Godin-Beekmann, S., Kivi, R., Stübi, R., and Zehner, C.: Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2, *Atmos. Meas. Tech.*, 8, 2093-2120, doi:10.5194/amt-8-2093-2015, 2015.
- [RD47] Keppens, A., Lambert, J.-C., Granville, J., Hubert, D., Verhoelst, T., Compernelle, S., Latter, B., Kerridge, B., Siddans, R., Boynard, A., Hadji-Lazarou, J., Clerbaux, C., Wespes, C., Hurtmans, D. R., Coheur, P.-F., van Peet, J. C. A., van der A, R. J., Garane, K., Koukouli, M. E., Balis, D. S., Delcloo, A., Kivi, R., Stübi, R., Godin-Beekmann, S., Van Roozendael, M., and Zehner, C.: Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 2: Ground-based validation of nadir ozone profile data products, *Atmos. Meas. Tech.*, 11, 3769–3800, doi:10.5194/amt-11-3769-2018, 2018.



- [RD48] Keppens, A., Compernelle, S., Verhoelst, T., Hubert, D., and Lambert, J.-C.: Harmonization and comparison of vertically resolved atmospheric state observations: methods, effects, and uncertainty budget, *Atmos. Meas. Tech.*, 12, 4379–4391, doi:10.5194/amt-12-4379-2019, 2019.
- [RD49] Kerr, J. B., I. A. Asbridge, and W. F. J. Evans, Intercomparison of total ozone measured by the Brewer and Dobson spectrophotometers at Toronto, *J. Geophys. Res.*, 93 (D9), 11129–11140, 1988.
- [RD50] Kiefer, M., T. von Clarmann, U. Grabowski, M. De Laurentis, R. Mantovani, M. Milz, and M. Ridolfi. Characterization of MIPAS elevation pointing. *Atmos. Chem. Phys.*, 7 (6): 1615–1628, doi: 10.5194/acp-7-1615-2007, 2007.
- [RD51] Komhyr, W. D., Mateer, C. L., and Hudson, R. D.: Effective Bass-Paur 1985 ozone absorption coefficients for use with Dobson ozone spectrophotometers, *J. Geophys. Res.*, 98, 20451–20465, <https://doi.org/10.1029/93JD00602>, 1993.
- [RD52] Koukouli, M. E., Balis, D. S., Loyola, D., Valks, P., Zimmer, W., Hao, N., Lambert, J.-C., Van Roozendaal, M., Lerot, C. and Spurr, R. J. D., Evaluating a new homogeneous total ozone climate data record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, *J. Geophys. Res. Atmos.*, 120, 12296–12312, doi:10.1002/2015JD023699, 2015.
- [RD53] Koukouli, M. E., Zara, M., Lerot, C., Fragkos, K., Balis, D., van Roozendaal, M., Allaart, M. A. F., and van der A, R. J.: The impact of the ozone effective temperature on satellite validation using the Dobson spectrophotometer network, *Atmos. Meas. Tech.*, 9, 2055–2065, <https://doi.org/10.5194/amt-9-2055-2016>, 2016.
- [RD54] Kramarova, N. A., Bhartia, P. K., Jaross, G., Moy, L., Xu, P., Chen, Z., DeLand, M., Froidevaux, L., Livesey, N., Degenstein, D., Bourassa, A., Walker, K. A., and Sheese, P.: Validation of ozone profile retrievals derived from the OMPS LP version 2.5 algorithm against correlative satellite measurements, *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2017-431, in review, 2017.
- [RD55] Laeng, A., D. Hubert, T. Verhoelst, T. von Clarmann, B.M. Dinelli, A. Dudhia, P. Raspollini, G. Stiller, U. Grabowski, A. Keppens, M. Kiefer, V. Sofieva, L. Froidevaux, K.A. Walker, J.-C. Lambert, C. Zehner, The ozone climate change initiative: Comparison of four Level-2 processors for the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), *Remote Sensing of Environment*, 162, 316-343, doi:10.1016/j.rse.2014.12.013, 2015.
- [RD56] Laeng, A., Eckert, E., von Clarmann, T., Kiefer, M., Hubert, D., Stiller, G., Grabowski, U., Glatthor, N., Plieninger, J., Kellmann, S., Linden, A., Lossow, S., Babenhauserheide, A., Froidevaux, L., and Walker, K.: On the improved stability of the version 7 MIPAS ozone record, *Atmos. Meas. Tech. Discuss.*, <https://doi.org/10.5194/amt-2017-345>, in review, 2017.
- [RD57] Lambert, J.-C., D. S. Balis, P. Gerard, J. Granville, Y. Livschitz, D. Loyola, R. Spurr, P. Valks, and M. Van Roozendaal, UPAS / GDOAS 4.0 Upgrade of the GOME Data Processor for Improved Total Ozone Columns – Delta Validation Report, Ed. by J.-C. Lambert (IASB) and D. Balis (AUTH), Tech. Note ERSE-CLVL-EOPG-TN-04-0001, European Space Agency, Frascati, Italy, 2004.
- [RD58] Lambert, J.-C., M. E. Koukouli, D. S. Balis, J. Granville, C. Lerot, D. Pieroux, and M. Van Roozendaal, GDP 5.0 - Upgrade of the GOME Data Processor for Improved Total Ozone Column - Validation Report for ERS-2 GOME GDP 5.0 Total Ozone Column, Edited by J.-C. Lambert (IASB) and M. E. Koukouli (AUTH), Tech. Note TN-IASB-GOME-GDP5-VR, Issue/Rev. 1/A, 55 pp., 6 May 2011.



- [RD59] Lambert, J.-C., C. De Clercq, and T. von Clarmann, Comparing and merging water vapour observations: A multi-dimensional perspective on smoothing and sampling issues, Chapter 9 (p. 177-199) of book “Monitoring Atmospheric Water Vapour: Ground-Based Remote Sensing and In-situ Methods”, N. Kämpfer (Ed.), ISSI Scientific Report Series, Vol. 10, Edition 1, 326 p., ISBN: 978-1-4614-3908-0, doi:10.1007/978-1-4614-3909-7_2, Springer, New York, 2012.
- [RD60] Langerock, B.; De Mazière, M.; Hendrick, F.; Vigouroux, C.; Desmet, F.; Dils, B. and Niemeijer, S., Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations, *Geosci. Model Dev.*, 8, 911-921, 2015.
- [RD61] Lerot, C., M. Van Roozendaal, R. Spurr, et al., Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat and GOME-2/Metop-A, *Journal of Geophysical Research*, 119, 1–20, doi:10.1002/2013JD020831, 2014.
- [RD62] Liu, G., D. W. Tarasick, V. E. Fioletov, C. E. Sioris, and Y. J. Rochon, Ozone correlation lengths and measurement uncertainties from analysis of historical ozonesonde data in North America and Europe, *J. Geophys. Res.*, 114, D04112, doi:10.1029/2008JD010576, 2009.
- [RD63] Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L., Millán Valle, L. F., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., and Martinez, E.: Aura Microwave Limb Sounder (MLS). Version 4.2x Level 2 data quality and description document, Tech. rep. JPL D-33509 Rev. C, Jet Propulsion Laboratory, http://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf, 2017.
- [RD64] Loyola, D.G., M. Coldewey-Egbers, M. Dameris, H. Garny, A. Stenke, M. van Roozendaal, C. Lerot, D. Balis, and M. Koukouli, Global long-term monitoring of the ozone layer – a prerequisite for predictions, *Int. J. Remote Sens.*, 30, 4295-4318, 2009.
- [RD65] McCormick, M. P., Lei, L., Hill, M. T., Anderson, J., Querel, R., and Steinbrecht, W.: Early results and validation of SAGE III-ISS ozone profile measurements from onboard the International Space Station, *Atmos. Meas. Tech.*, 13, 1287–1297, <https://doi.org/10.5194/amt-13-1287-2020>, 2020.
- [RD66] McLinden, C., V. Fioletov, C. Haley, N. Lloyd, C. Roth, D. Degenstein, A. Bourassa, C. McElroy, and E. Llewellyn. An evaluation of Odin/OSIRIS limb pointing and stratospheric ozone through comparisons with ozonesondes. *Can. J. Phys.*, 85 (11): 1125–1141, doi:10.1139/P07-112, 2007.
- [RD67] Mégie, G., J. Y. Allain, M. L. Chanin, and J. E. Blamont, Vertical profile of stratospheric ozone by lidar sounding from the ground, *Nature* 270, 329–331, doi:10.1038/270329a0, 1977.
- [RD68] Moy, L., Bhartia, P. K., Jaross, G., Loughman, R., Kramarova, N., Chen, Z., Taha, G., Chen, G., and Xu, P.: Altitude registration of limb-scattered radiation, *Atmos. Meas. Tech.*, 10, 167-178, doi:10.5194/amt-10-167-2017, 2017.
- [RD69] Nair, P. J., S. Godin-Beekmann, L. Froidevaux, L. E. Flynn, J. M. Zawodny, J. M. Russell III, A. Pazmiño, G. Ancellet, W. Steinbrecht, H. Claude, T. Leblanc, S. McDermid, J. A. E. van Gijsel, B. Johnson, A. Thomas, D. Hubert, J.-C. Lambert, H. Nakane, and D. P. J. Swart, Relative drifts and stability of satellite and ground-based stratospheric ozone profiles at NDACC lidar stations, *Atmospheric Measurement Techniques*, 5, 1301-1318 doi:10.5194/amt-5-1301-2012, 2012.
- [RD70] Ozone_cci+ Climate Assessment Report (CAR), <https://climate.esa.int/en/projects/ozone/key-documents>, 07/03/2012.
- [RD71] Ozone_cci Product Validation and Intercomparison Report (PVIR v2), Ozone_cci_Phase-II_PVIR_2.0, Issue 2.0 / Rev. 0, <https://climate.esa.int/en/projects/ozone/key-documents>, 30/06/2016.



- [RD72] Pommereau, J.P. and Goutail, F., O₃ and NO₂ ground-based measurements by visible spectrometry during arctic winter and spring 1988. *Geophysical Research Letters*, 15, doi:10.1029/88GL00479, 1988.
- [RD73] Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding*, World Scientific Publishing, Singapore, p. 238, 2000.
- [RD74] Rong, P. P., J. M. Russell III, M. G. Mlynczak, E. E. Remsberg, B. T. Marshall, L. L. Gordley, and M. López-Puertas, Validation of Thermosphere Ionosphere Mesosphere Energetics and Dynamics/Sounding of the Atmosphere using Broadband Emission Radiometry (TIMED/SABER) v1.07 ozone at 9.6 μm in altitude range 15–70 km, *J. Geophys. Res.*, 114, D04306, doi:10.1029/2008JD010073, 2009.
- [RD75] Sentinel-5 precursor / TROPOMI Level 2 Algorithm Theoretical Basis Document O3 Total Column, DLR/BIRA-IASB, S5P-L2-DLR-ATBD-400A, Issue 2.33, 62pp., <https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-Total-Ozone,04/06/2021>.
- [RD76] Sheese, P. E., Walker, K. A., Boone, C. D., Bernath, P. F., Froidevaux, L., Funke, B., Raspollini, P., and von Clarmann, T.: ACE-FTS ozone, water vapour, nitrous oxide, nitric acid, and carbon monoxide profile comparisons with MIPAS and MLS, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 186, 63-80, doi:10.1016/j.jqsrt.2016.06.026, 2017.
- [RD77] Sheese, P. E., Walker, K. A., Boone, C. D., Bourassa, A. E., Degenstein, D. A., Froidevaux, L., McElroy, C. T., Murtagh, D., Russell III, J. M., and Zou, J.: Assessment of the quality of ACE-FTS stratospheric ozone data, *Atmos. Meas. Tech.*, 15, 1233–1249, <https://doi.org/10.5194/amt-15-1233-2022>, 2022.
- [RD78] Simmons, A. J., Poli, P., Dee, D. P., Berrisford, P., Hersbach, H., Kobayashi, S. and Peubey, C., Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q.J.R. Meteorol. Soc.*, 140: 329–353. doi: 10.1002/qj.2317, 2014.
- [RD79] Smit, H., and the Panel for the Assessment of Standard Operating Procedures for Ozonesondes (ASOPOS), Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, GAW Report No. 201, September 2011.
- [RD80] Sofieva, V. F., N. Rahpoe, J. Tamminen, E. Kyrölä, N. Kalakoski, M. Weber, A. Rozanov, C. von Savigny, A. Laeng, T. von Clarmann, G. Stiller, S. Lossow, D. Degenstein, A. Bourassa, C. Adams, C. Roth, N. Lloyd, P. Bernath, R. J. Hargreaves, J. Urban, D. Murtagh, A. Hauchecorne, F. Dalaudier, M. van Roozendaal, N. Kalb, and C. Zehner. Harmonized dataset of ozone profiles from satellite limb and occultation measurements. *Earth System Science Data*, 5 (2): 349–363, doi: 10.5194/essd-5-349-2013, 2013.
- [RD81] Sofieva, V. F., Kyrölä, E., Laine, M., Tamminen, J., Degenstein, D., Bourassa, A., Roth, C., Zawada, D., Weber, M., Rozanov, A., Rahpoe, N., Stiller, G., Laeng, A., von Clarmann, T., Walker, K. A., Sheese, P., Hubert, D., van Roozendaal, M., Zehner, C., Damadeo, R., Zawodny, J., Kramarova, N., and Bhartia, P. K.: Merged SAGE II, Ozone_cci and OMPS ozone profile dataset and evaluation of ozone trends in the stratosphere, *Atmos. Chem. Phys.*, 17, 12533-12552, doi:10.5194/acp-17-12533-2017, 2017.
- [RD82] Staehelin J., J. Kerr, R. Evans and K. Vanicek, Comparison of total ozone measurements of Dobson and Brewer spectrophotometers and recommended transfer functions, WMO TD N. 1147, No 149, 2003.



- [RD83] Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Tarasick, D. W., Van Malderen, R., Smit, H. G. J., Vömel, H., Morris, G. A., Johnson, B. J., Cullis, P. D., Rene Stübi, R., Davies, J. and Yan, M. M., An Examination of the Recent Stability of Ozonesonde Global Network Data, Earth and Space Science Open Archive, preprint, <https://doi.org/10.1002/essoar.10511590.1>, 2022.
- [RD84] Thompson, A.M., J.C. Witte, R.D. McPeters, S.J. Oltmans, F.J. Schmidlin, J.A. Logan, M. Fujiwara, V.W.J.H. Kirchhoff, F. Posny, G.J.R. Coetzee, B. Hoegger, S. Kawakami, T. Ogawa, B.J. Johnson, H. Vömel and G. Labow, Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998-2000 Tropical ozone climatology 1. Comparison with Total Ozone Mapping Spectrometer (TOMS) and ground-based measurements, *J. Geophys. Res.*, Vol. 108 No. D2, 8238, doi: 10.1029/2001JD000967, 30 January 2003.
- [RD85] Thompson, A.M., J.C. Witte, S.J. Oltmans, F.J. Schmidlin, J.A. Logan, M. Fujiwara, V.W.J.H. Kirchhoff, F. Posny, G.J.R. Coetzee, B. Hoegger, S. Kawakami, T. Ogawa, J.P.F. Fortuin, and H.M. Kelder, Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998-2000 Tropical ozone climatology 2. Tropospheric variability and the zonal wave-one, *J. Geophys. Res.*, Vol. 108 No. D2, 8241, doi: 10.1029/2002JD002241, 31 January 2003.
- [RD86] Tukiainen, S., Kyrölä, E., Tamminen, J., Kujanpää, J., and Blanot, L.: GOMOS bright limb ozone data set, *Atmos. Meas. Tech.*, 8, 3107-3115, doi:10.5194/amt-8-3107-2015, 2015.
- [RD87] Vandenbussche, S., C. De Clercq, J.-C. Lambert, R. Spurr, and T. von Clarmann. EC FP6 GEOmon Technical note D4.2.1 - Multi-dimensional characterisation of remotely sensed data - Chapter 4: Satellite measurements of limb-scattered ultraviolet-visible light. GEOmon TN-IASB-OBSOP / Chapter 4, BIRA-IASB, July 2010.
- [RD88] Vandenbussche, S., J.-C. Lambert, and R. Spurr, EC FP6 GEOmon Technical note D4.2.1 - Multi-dimensional characterisation of remotely sensed data - Chapter 5: Satellite measurements of solar/stellar occultation. GEOmon TN-IASB-OBSOP / Chapter 5, BIRA-IASB, December 2010.
- [RD89] van Roozendaal, M., P. Peters, H. K. Roscoe, et al., Validation of Ground-Based Visible Measurements of Total Ozone by Comparison with Dobson and Brewer Spectrophotometers, *J. Atmos. Chem*, 29, 55–83, 1998.
- [RD90] von Clarmann, T., C. De Clercq, M. Ridolfi, M. Höpfner, and J.-C. Lambert, The horizontal resolution of MIPAS, *Atmos. Meas. Tech.*, Vol. 2, 47-54, 2009.
- [RD91] von Savigny, C., H. Bovensmann, K. Bramstedt, S. Dikty, F. Ebojje, A. Jones, S. Noël, A. Rozanov, and B.-M. Sinnhuber. Indications for long-term trends and seasonal variations in the SCIAMACHY Level 1 version 6.03 tangent height information. Technical Report TN-IUP-scia-pointing-2009-01, IUP Bremen, June 2009.
- [RD92] Wang, H. J. R., Damadeo, R., Flittner, D., Kramarova, N., Taha, G., Davis, S., et al. (2020). Validation of SAGE III/ISS solar occultation ozone products with correlative satellite and ground based measurements. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032430. <https://doi.org/10.1029/2020JD032430>, 2020.
- [RD93] Waters, J. W. et al., "The Earth observing system microwave limb sounder (EOS MLS) on the aura Satellite," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 5, pp. 1075-1092, doi:10.1109/TGRS.2006.873771, 2006.
- [RD94] World Meteorological Organization, Scientific assessment of ozone depletion: 2006, Global Ozone Res. Monit. Proj., Rep. 50, Geneva, Switzerland, 2006.



9 Terms and definitions

9.1 Terminology

In Table 9.1, terms and definitions as recommended by CEOS WGCV and by standards development organisations of international recognition have been transcript from reference documents [RD10] to [RD18]. In some cases, terms and definitions peculiar to forecast systems are also proposed. They are expected to evolve as these organisations regularly update their standards and as further standardisation and harmonisation occur.

Table 9.1 - Recommended terms and definitions.

TERM	DEFINITION	SOURCE
accuracy	closeness of agreement between a quantity value obtained by measurement and the true value of the measurand; note that <u>it is not a quantity</u> and it is not given a numerical quantity value	VIM, GUM
area (volume) of representativeness	the area (volume) in which the concentration does not differ from the concentration at the station by more than a specific range	Larssen
bias	(1) systematic error of indication of a measuring system (2) estimate of a systematic measurement error (3) estimate of a systematic forecast error	VIM VIM GAS
calibration	(1) the process of quantitatively defining the system responses to known, controlled signal inputs (2) operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication	CEOS VIM
dead band (or neutral zone)	maximum interval through which a value of a quantity being measured can be changed in both directions without producing a detectable change in the corresponding indication	VIM
detection limit	measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component, given a probability α of falsely claiming its presence	VIM
error	(1) measured quantity value minus a reference quantity value (2) difference of quantity value obtained by measurement and true value of the measurand (3) difference of forecast value and a, estimate of the true value	VIM CEOS
establish	define, document and implement	CDRH
field-of-regard	an area of the object space scanned by the field-of-view of a scanning sensor	NIST
field-of-view	the solid angle from which the detector receives radiation	NIST
footprint	the area of a target encircled by the field-of-view of a detector of radiation, or irradiated by an active system	NIST



influence quantity	quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result	VIM
in situ measurement	(1) a direct measurement of the measurand in its original place (2) any sub-orbital measurement of the measurand	GEOSS
measurand	quantity intended to be measured	VIM
metadata	data about the data; parameters that describe, characterise, and/or index the data	WMO
monitoring	(1) systematic evaluation over time of some quantity (2) by extension, evaluation over time of the performance of a system, of the occurrence of an event etc.	NIST
point-to-area (point-to-volume) representativeness	the probability that a point measurement lies within a specific range of area-average (volume-average) concentration value	Nappo
precision	closeness of agreement between quantity values obtained by replicate measurements of a quantity on the same or similar object under specified conditions	VIM
process validation	establishing documented evidence of a high degree of assurance that a specific process will consistently produce a product meeting its pre-determined specifications and quality characteristics	CDRH
quality assessment (QA)	QA refers to the overall management of the processes involved in obtaining the data	CEOS
quality control (QC)	QC refers to the activities undertaken to check and optimise accuracy and precision of the data after its collection	CEOS
quality indicator (QI)	a means of providing a user of data or derived product with sufficient information to assess its suitability for a particular application. This information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally SI), but can be presented as a numeric or a text descriptor, provided the quantitative linkage is defined.	QA4EO
radiometric calibration	a determination of radiometric instrument performance in the spatial, spectral, and temporal domains in a series of measurements, in which its output is related to the true value of the measured radiometric quantity	NIST
random error	(1) component of measurement error that in replicate measurements varies in an unpredictable manner; note that random measurement error equals measurement error minus systematic measurement error (2) component of forecast error that varies in an unpredictable manner	VIM
relative standard uncertainty	standard measurement uncertainty divided by the absolute value of the measured quantity value	VIM
repeatability	measurement precision under set of conditions including the same measurement procedure, same operator, same measuring system, same operating conditions and same location, and replicated measurements over a short period of time	VIM



representativeness	the extent to which a set of measurements taken in a given space-time domain reflect the actual conditions in the same or different space-time domain taken on a scale appropriate for a specific application	Nappo
reproducibility	measurement precision under a set of conditions including different locations, operators, and measuring systems	VIM
resolution	(1) the least angular/linear/temporal/spectral distance between two identical point sources of radiation that can be distinguished according to a given criterion (2) the least vertical/geographical/temporal distance between two identical atmospheric features that can be distinguished in a gridded numerical product or in time series of measurements; resolution is equal to or coarser than vertical/geographical/temporal sampling of the grid or the measurement time series	NIST
stability	ability of a measuring system to maintain its metrological characteristics constant with time	VIM
systematic error	component of measurement error that in replicate measurements remains constant or varies in a predictable manner	VIM
traceability	property of a measurement result relating the result to a stated metrological reference (free definition and not necessarily SI) through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty	VIM
tropopause	the region of the atmosphere where the environmental temperature lapse rate changes from positive (in the troposphere) to negative (in the stratosphere) the lowest level at which the lapse rate decreases to 2 °C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C/km occasionally, a second tropopause may be found if the lapse rate above the first tropopause exceeds 3 °C/km	WMO
uncertainty	non-negative parameter that characterizes the dispersion of the quantity values that are being attributed to a measurand, based on the information used	VIM
validation	(1) the process of assessing, by independent means, the quality of the data products derived from the system outputs (2) verification where the specified requirements are adequate for an intended use (3) the process of assessing, by independent means, the degree of correspondence between the value of the radiometric quantity derived from the output signal of a calibrated radiometric device and the actual value of this quantity. (4) confirmation by examination and provision of objective evidence that specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled	CEOS VIM NIST CDRH



verification	(1) the provision of objective evidence that a given data product fulfils specified requirements; note that, when applicable, measurement uncertainty should be taken into consideration. (2) the provision of objective evidence that the design outputs of a particular phase of the software development life cycle meet all of the specified requirements for that phase	VIM CDRH
vicarious calibration	a post-launch radiometric calibration of sensors performed with the use of natural or artificial sites or objects on the surface of the Earth (as opposed to calibration techniques using onboard standards such as lamps, blackbodies, solar diffuse reflecting panels etc.)	NIST



9.2 Abbreviations and acronyms

Note of best practice: Using an acronym is acceptable if it has been defined the first time it appears in a document. The same applies to chemical abbreviations. In documents targeting a wide spectrum of potential readers, like user manuals and validation reports, it is recommended to avoid systematic use of acronyms and abbreviations except for those with frequent occurrence, and those widely understood by the general public. For example, acronyms such as CFCs and ESA are acceptable. Acronyms such as ECSS and ICTT-QMF are not. Before using acronyms and abbreviations, authors should keep in mind that it is annoying and difficult – especially in Web-based documents unless the acronyms are available as hyperlinks – to turn over several pages in a document to verify the meaning.

AK	Averaging Kernel
AMF	Air Mass Factor or optical enhancement factor
ATBD	Algorithm Theoretical Basis Document
AUTH	Aristotle University of Thessaloniki
BIRA-IASB	Belgian Institute for Space Aeronomy
C3S	Copernicus Climate Change Service
CCI	ESA's Climate Change Initiative programme
CEOS	Committee on Earth Observation Satellites
CMUG	Climate Modelling User Group of the CCI programme
CDRP	Climate Research Data Package
CRG	Climate Research Group of the Ozone_cci+ project
DARD	Data Access Requirement Document
DFS	Degree of Freedom of the System
DHF	Data Host Facility
DIAL	Differential Absorption LIDAR
DLR	German Aerospace Centre
DOAS	Differential Absorption Optical Spectroscopy
DU	Dobson Unit – unit of vertical column density (2.69 10 ¹⁶ molec.cm ⁻²)
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
ECSS	European Corporation for Space Standardization
Envisat	ESA's Environmental Satellite, launched March 1, 2002
EO	Earth Observation
EOST	Earth Observation Science Teams of the Ozone_cci+ project
EPS	EUMETSAT Polar System
ERA-I	ECMWF ReAnalysis Interim
ERA5	ECMWF ReAnalysis 5
ERS-2	ESA's Earth Remote Sensing satellite 2, launched April 21, 1995
ESA	European Space Agency
ESRIN	European Space Research Institute
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FMI	Finnish Meteorological Institute
FRM	Fiducial Reference Measurements
FTIR	Fourier Transform Infra-Red spectrometer
GAS	GMES Atmospheric Service
GAW	WMO's Global Atmosphere Watch
GCOS	Global Climate Observing System
GDP	GOME Data Processor
GEO	Group on Earth Observation



GEOSS	Global Earth Observation System of Systems
GMES	Global Monitoring for Environment and Security
GOME	Global Ozone Monitoring Experiment
GOMOS	Global Ozone Monitoring by Occultation of Stars
GPS	Global Positioning System
GUM	Guide to the expression of uncertainty in a measurement
HALOE	Halogen Occultation Experiment
ICTT-QMF	Inter-Commission Task Team on Quality Management Framework
IGACO	Integrated Global Atmospheric Chemistry Observation strategy
IGOS	Integrated Global Observation Strategy
INSPIRE	Infrastructure for Spatial Information in the European Community
IPF	Instrument Processing Facility
I/O tools	Input/Output tools
IR	INSPIRE Implementation Rule
ISO	International Organization for Standardization
ISSI	International Space Science Institute
JCGM	Joint Committee for Guides in Metrology
KNMI	Royal Dutch Meteorological Institute
lidar	light detection and ranging
LP	Limb Profile
MetOp	EUMETSAT's Meteorological Operational satellite
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MIPAS FR	MIPAS operated at Full (nominal) Resolution
MIPAS RR	MIPAS operated at Reduced (optimised) Resolution
MLS	Microwave Limb Sounder
Multi-TASTE	Technical ASsistance To the multi-mission validation of Envisat and Third Party Missions using spectrometers, radiometers and sondes
MWR	MicroWave Radiometer
NCEP	National Centers for Environmental Prediction
NDACC	Network for the Detection of Atmospheric Composition Change
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NP	Nadir Profile
O3	Ozone
OE	Optimal Estimation
OMI	Ozone Monitoring Instrument
PSD	Product Specification Document
PVP	Product Validation Plan
QA4EO	Quality Assurance framework for Earth Observation
RAL	Rutherford Appleton Laboratory
S5P	Sentinel-5 Precursor
SAGE	Stratospheric Aerosol and Gas Experiment
SBUV	Solar Backscatter Ultraviolet
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartography
SGP	SCIAMACHY Ground Processor
SH	Southern Hemisphere
SHADOZ	Southern Hemisphere ADditional Ozonesondes
SNPP	Suomi National Polar-orbiting Partnership satellite
SZA	Solar Zenith Angle



TBD	To Be Determined
TEMIS	Tropospheric Emission Monitoring Internet Service
TOC	Total Ozone Column
TOMS	Total Ozone Mapping Spectrometer
TROPOMI	TROPOspheric Monitoring Instrument
UARS	Upper Atmosphere Research Satellite, launched September 15, 1991
ULB	Université Libre de Bruxelles
URD	User Requirement Document
USM	Upper Stratosphere/Mesosphere
UT	Upper Troposphere
UTLS	Upper Troposphere/Lower Stratosphere
UVVIS	DOAS UV-visible spectrometer (generic)
VALT	Validation team of the Ozone_cci+ project
VIM	International Vocabulary of Metrology – Basic and general concepts and associated terms
VMR	Volume Mixing Ratio
WGCV	CEOS Working Group on Calibration and Validation
WMO	World Meteorological Organization
WOUDC	World Ozone and Ultraviolet Radiation Data Center

END OF DOCUMENT