

Algorithm Theoretical Baseline Document: Level 3 gridded data

Version 4.3

18 October 2021

ROM SAF Consortium
Danish Meteorological Institute (DMI)
European Centre for Medium-Range Weather Forecasts (ECMWF)
Institut d'Estudis Espacials de Catalunya (IEEC)
Met Office (UKMO)

DOCUMENT AUTHOR TABLE

	Author(s)	Function	Date
Prepared by:	Hans Gleisner	ROM SAF Climate Coordinator	18/10 2021
Reviewed by (internal):	Johannes K. Nielsen	ROM SAF Scientist	21/2 2020
Approved by:	Kent B. Lauritsen	ROM SAF Project Manager	18/10 2021

DOCUMENT CHANGE RECORD

Version	Date	By	Description
1.0	11/11 2009	HGL	Original version for PCR2 Review.
2.0	18/02 2011	HGL	Revised version for PCR2 Close Out.
2.1	13/04 2011	HGL	Revised according to Action 1 from PCR2 Close Out.
2.2	18/03 2013	HGL	Release for PCR5 Review. RIDs 90 and 91 from ORR2 implemented. Added algorithms for offline Level 3 <i>dry</i> products: (1) now addressing dry variables, (2) more complete description of the QC screening procedures, (3) list of Reference Documents revised, (4) Annex II removed.
2.3	10/06 2013	HGL	Revised version after the PCR5 Review: closing all ATBD RIDs 003 to 035.
2.4	26/02 2016	HGL	Revision for PCR-RE1 Review: New Section 7 and revised Section 4. Extended Table1, listing all planned Level 3 products.
2.5	14/04 2016	HGL	Revision following RIDs from PCR-RE1 reviewers: RIDs 002, 013, 014, 036, 037 implemented.
2.6	28/04 2016	HGL	Revision following the PCR-RE1 Review Meeting. RID 003 implemented (Section 4.1, Eq. 6 and related text).
3.0	8/6 2018	HGL	Version for DRR-RE1 and ORR3/5 reviews. List of changes: Section 5.1: Editorials. Section 5.2.1: Changes made to the QC procedures, particularly QC-2. Section 5.2.3: Included a generalization of the sub-binning technique (Eqs. 24-26). Added the weighted standard deviation (Eq. 27). Section 5.2.4: Added an equation (Eq. 31) for the uncertainty, related to the generalization of the sub-binning technique. Extended description of sampling errors (new Eq. 34) Section 5.3: New section describing generation of time series anomaly data (Eqs. 41-48).

Version	Date	By	Description
			Section 6.4: New examples, new figures (Metop instead of COSMIC-1).
3.1	31/8 2018	HGL	Revised version after DRR-RE1 and ORR3/5 reviews: RIDs 149,151,152,156,157,201,202: Minor changes implemented. RIDs 115-118,120,124-126,129,136-141,143-148,150-155,158-163,166-170,197-200,203: Editorial changes implemented.
3.2	25/11 2018	HGL	Updated version based on ICDR concept discussions at ROM SAF SG22: Page 4: Update of the ROM SAF introduction. Section 1.1: Tables 1 and 2 updated. New Table 3. Chapter 7: Text was copied from Chapter 7 to Chapter 6, and Chapter 7 was then removed.
4.0	21/2 2020	HGL	Updated version for the EPS-SG PDCR review: Section 1.1: New subsection 1.1.1 and Table 4. Updated version for the ORR12 review: Section 1.1: Addition of Metop-C to the list of gridded offline products in Table 3. Section 4.1: Update of text to reflect the introduction of ERA5.1 reanalysis data, which replaces ERA-Interim.
4.1	14/5 2020	HGL	Updated version after ORR12 review implementing the following changes and ORR12 RIDs: Section 1.1: text and table about future EPS-SG Day 1 products removed [RIDs 043, 098]. Section 1.1.1: this section about future EPS-SG Day 1 products removed [RIDs 043, 098]. Section 3: editorial change [RID 036]. Section 4.1: text and table updated [RIDs 037, 041]. Section 4.3: text updated [RID 040].
4.2	28/9 2021	HGL	Updated version submitted for the PCR review for Sentinel-6 NTC products implementing the following changes: Section 1.1: Introduced NTC data in first paragraph of text and in Table 1.4. Section 1.3: Added NTC to the acronym list. Section 1.4: Added NTC to product types, and introduced description of general terms. Section 2: Sentinel-6 mentioned in third paragraph.
4.3	18/10 2021	HGL	Updated version prepared during PCR review process implementing the following: - Sec. 1.4: Editorial corrections - Editorial corrections (Level 1b/1B, NTC)

ROM SAF

The Radio Occultation Meteorology Satellite Application Facility (ROM SAF) is a decentralised processing centre under EUMETSAT which is responsible for operational processing of GRAS radio occultation (RO) data from the Metop and Metop-SG satellites and radio occultation data from other missions. The ROM SAF delivers bending angle, refractivity, temperature, pressure, humidity, and other geophysical variables in near real-time for NWP users, as well as reprocessed Climate Data Records (CDRs) and Interim Climate Data Records (ICDRs) for users requiring a higher degree of homogeneity of the RO data sets. The CDRs and ICDRs are further processed into globally gridded monthly-mean data for use in climate monitoring and climate science applications.

The ROM SAF also maintains the Radio Occultation Processing Package (ROPP) which contains software modules that aid users wishing to process, quality-control and assimilate radio occultation data from any radio occultation mission into NWP and other models.

The ROM SAF Leading Entity is the Danish Meteorological Institute (DMI), with Cooperating Entities: i) European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading, United Kingdom, ii) Institut D'Estudis Espacials de Catalunya (IEEC) in Barcelona, Spain, and iii) Met Office in Exeter, United Kingdom. To get access to our products or to read more about the ROM SAF please go to: <http://www.romsaf.org>

Intellectual Property Rights

All intellectual property rights of the ROM SAF products belong to EUMETSAT. The use of these products is granted to every interested user, free of charge. If you wish to use these products, EUMETSAT's copyright credit must be shown by displaying the words “copyright (year) EUMETSAT” on each of the products used.

List of Contents

1. INTRODUCTION.....	6
1.1 PURPOSE OF DOCUMENT	6
1.2 APPLICABLE AND REFERENCE DOCUMENTS.....	9
1.2.1 <i>Applicable documents</i>	9
1.2.2 <i>Reference documents</i>	10
1.3 ACRONYMS AND ABBREVIATIONS.....	12
1.4 DEFINITIONS	13
2. THE RADIO OCCULTATION TECHNIQUE	15
3. HEIGHT VARIABLES	16
4. LEVEL 1B AND LEVEL 2 PROFILE DATA	18
4.1 FROM MEASUREMENTS TO ATMOSPHERIC PROFILES.....	18
4.2 ATMOSPHERIC PROFILE DATA	21
4.3 TIME AND LOCATION OF ATMOSPHERIC PROFILES	24
4.4 UNCERTAINTY OF ATMOSPHERIC PROFILES.....	25
4.5 SPATIAL AND TEMPORAL SAMPLING OF THE ATMOSPHERE	31
4.6 LONG-TERM STABILITY OF DATA	34
5. LEVEL 3 ALGORITHMS.....	35
5.1 ALGORITHMS OVERVIEW.....	35
5.2 FROM PROFILES TO GRIDDED MONTHLY MEAN DATA	36
5.2.1 <i>Quality control of profiles</i>	36
5.2.2 <i>Interpolation of profiles onto the Level 3 height grids</i>	38
5.2.3 <i>Averaging in grid boxes</i>	39
5.2.4 <i>Estimating uncertainty in the means</i>	41
5.2.5 <i>Estimating a priori information in the means</i>	43
5.3 TIME SERIES AND ANOMALY DATA.....	45
6. LEVEL 3 GRIDDED DATA PRODUCTS	47
6.1 PRODUCT TYPES: OFFLINE DATA, CDRs, AND ICDRS	47
6.2 GEOPHYSICAL VARIABLES AND HEIGHT VARIABLES	47
6.3 DATA GRIDS.....	47
6.4 NUMERICAL DATA.....	48
6.5 EXAMPLE DATA	48
ANNEX I: MONITORING LONG-TERM STABILITY OF THE CLIMATE DATA.....	67

1. Introduction

1.1 Purpose of document

The purpose of this document is to provide a description of the assumptions, methods, and algorithms that are used to generate ROM SAF Level 3 gridded monthly mean data products. The data products covered by this ATBD are listed in Tables 1.1 to 1.4. They include *Climate data Records (CDRs)*, *Interim Climate Data Records (ICDRs)*, as well as *Offline* and *Non Time Critical (NTC) data products*. Note that the listed data products include, or may include, products in development as well as products with operational status. The status of all ROM SAF data products is available at the website: <http://www.romsaf.org>.

The product requirements baseline is the PRD, version 2.3 [AD.3]. The methods and algorithms described in the present ATBD have been implemented in the *romclim* software package, which is included in the *GPAC* processing system.

Section 2 provides an introductory background, while Section 3 gives an introduction to the different height variables that are relevant in conjunction with the RO technique. In Section 4 the profile input data to the Level 3 processing are described. The section gives an overview of the processing from raw measurements to atmospheric profiles, defines some of the key variables, and discusses the spatial and temporal sampling of the atmosphere. The quality and long-term stability of the data is also discussed. Section 5 describes the Level 3 algorithms in some detail, with examples of the resulting Level 3 gridded data given in Section 6.

Ref: SAF/ROM/DMI/ALG/GRD/001 Version: 4.3 Date: 18 October 2021	ATBD Level 3 gridded data	
---	------------------------------	---

Table 1.1. List of *Climate Data Records* covered by this ATBD.

Product ID	Product name	Product acronym	Satellite input	Format
GRM-28-L3-B-R1	Reprocessed bending angle grid	RBGMUL	Multi-satellite record: CHAMP, GRACE COSMIC-1, Metop <i>Level 1A data from EUM Secretariat and from CDAAC</i>	netCDF
GRM-28-L3-R-R1	Reprocessed refractivity grid	RRGMUL	"	"
GRM-28-L3-D-R1	Reprocessed dry temperature grid	RDGMUL	"	"
GRM-28-L3-Y-R1	Reprocessed dry pressure grid	RYGMUL	"	"
GRM-28-L3-Z-R1	Reprocessed dry geopotential height grid	RZGMUL	"	"
GRM-28-L3-T-R1	Reprocessed temperature grid	RTGMUL	"	"
GRM-28-L3-H-R1	Reprocessed specific humidity grid	RHGMUL	"	"
GRM-28-L3-C-R1	Reprocessed tropopause height grid	RCGMUL	"	"
GRM-29-L3-B-R1	Reprocessed bending angle grid	RBGMET	Metop <i>Level 1A data from EUM Secretariat</i>	netCDF
GRM-29-L3-R-R1	Reprocessed refractivity grid	RRGMET	"	"
GRM-29-L3-D-R1	Reprocessed dry temperature grid	RDGMET	"	"
GRM-29-L3-Y-R1	Reprocessed dry pressure grid	RYGMET	"	"
GRM-29-L3-Z-R1	Reprocessed dry geopotential height grid	RZGMET	"	"
GRM-29-L3-T-R1	Reprocessed temperature grid	RTGMET	"	"
GRM-29-L3-H-R1	Reprocessed specific humidity grid	RHGMET	"	"
GRM-29-L3-C-R1	Reprocessed tropopause height grid	RCGMET	"	"
GRM-30-L3-B-R1	Reprocessed bending angle grid	RBGCO1	COSMIC <i>Level 1A data from CDAAC</i>	netCDF
GRM-30-L3-R-R1	Reprocessed refractivity grid	RRGCO1	"	"
GRM-30-L3-D-R1	Reprocessed dry temperature grid	RDGCO1	"	"
GRM-30-L3-Y-R1	Reprocessed dry pressure grid	RYGCO1	"	"
GRM-30-L3-Z-R1	Reprocessed dry geopotential height grid	RZGCO1	"	"
GRM-30-L3-T-R1	Reprocessed temperature grid	RTGCO1	"	"
GRM-30-L3-H-R1	Reprocessed specific humidity grid	RHGCO1	"	"
GRM-30-L3-C-R1	Reprocessed tropopause height grid	RCGCO1	"	"
GRM-32-L3-B-R1	Reprocessed bending angle grid	RBGCHA	CHAMP <i>Level 1A data from CDAAC</i>	netCDF
GRM-32-L3-R-R1	Reprocessed refractivity grid	RRGCHA	"	"
GRM-32-L3-D-R1	Reprocessed dry temperature grid	RDGCHA	"	"
GRM-32-L3-Y-R1	Reprocessed dry pressure grid	RYGCHA	"	"
GRM-32-L3-Z-R1	Reprocessed dry geopotential height grid	RZGCHA	"	"
GRM-32-L3-T-R1	Reprocessed temperature grid	RTGCHA	"	"
GRM-32-L3-H-R1	Reprocessed specific humidity grid	RHGCHA	"	"
GRM-32-L3-C-R1	Reprocessed tropopause height grid	RCGCHA	"	"

Ref: SAF/ROM/DMI/ALG/GRD/001 Version: 4.3 Date: 18 October 2021	ATBD Level 3 gridded data	
---	------------------------------	---

Product ID	Product name	Product acronym	Satellite input	Format
GRM-33-L3-B-R1	Reprocessed bending angle grid	RBGGRA	GRACE <i>Level 1A data from CDAAC</i>	netCDF
GRM-33-L3-R-R1	Reprocessed refractivity grid	RRGGRA	"	"
GRM-32-L3-D-R1	Reprocessed dry temperature grid	RDGGRA	"	"
GRM-32-L3-Y-R1	Reprocessed dry pressure grid	RYGGRA	"	"
GRM-32-L3-Z-R1	Reprocessed dry geopotential height grid	RZGGRA	"	"
GRM-32-L3-T-R1	Reprocessed temperature grid	RTGGRA	"	"
GRM-32-L3-H-R1	Reprocessed specific humidity grid	RHGGRA	"	"
GRM-32-L3-C-R1	Reprocessed tropopause height grid	RCGGRA	"	"

Table 1.2. List of *Interim Climate Data Records* covered by this ATBD. The Level 1A input data to the ROM SAF processing is obtained from the EUMETSAT Secretariat.

Product ID	Product Name	Product acronym	Satellite input	File format
GRM-29-L3-B-I1	ICDR bending angle grid	IBGMET	Metop	netCDF
GRM-29-L3-R-I3	ICDR refractivity grid	IRGMET	"	"
GRM-29-L3-D-I3	ICDR dry temperature grid	ITGMET	"	"
GRM-29-L3-Y-I3	ICDR dry pressure grid	IHGMET	"	"
GRM-29-L3-Z-I3	ICDR dry geopotential height grid	IZGMET	"	"
GRM-29-L3-T-I3	ICDR temperature grid	IDGMET	"	"
GRM-29-L3-H-I3	ICDR specific humidity grid	IYGMET	"	"
GRM-29-L3-C-R1	ICDR tropopause height grid	ICGMET	"	"

Table 1.3. List of *Offline data products* covered by this ATBD. The Level 1A input data to the ROM SAF processing is obtained from the EUMETSAT Secretariat.

Product ID	Product Name	Product acronym	Satellite input	File format
GRM-93	Offline bending angle grid	OBGMEA	Metop-A	netCDF
GRM-94	Offline refractivity grid	ORGMEA	"	"
GRM-95	Offline temperature grid	OTGMEA	"	"
GRM-96	Offline specific humidity grid	OHGMEA	"	"
GRM-97	Offline dry geopotential height grid	OZGMEA	"	"
GRM-98	Offline dry temperature grid	ODGMEA	"	"
GRM-99	Offline dry pressure grid	OYGMEA	"	"
GRM-191	Offline tropopause height grid	OCGMEA	"	"
GRM-53	Offline bending angle grid	OBGMEB	Metop-B	netCDF
GRM-54	Offline refractivity grid	ORGMEB	"	"
GRM-55	Offline temperature grid	OTGMEB	"	"

Product ID	Product Name	Product acronym	Satellite input	File format
GRM-56	Offline specific humidity grid	OHGMEB	"	"
GRM-57	Offline dry geopotential height grid	OZGMEB	"	"
GRM-58	Offline dry temperature grid	ODGMEB	"	"
GRM-59	Offline dry pressure grid	OYGMEB	"	"
GRM-192	Offline tropopause height grid	OCGMEB	"	"
GRM-73	Offline bending angle grid	OBGMEC	Metop-C	netCDF
GRM-74	Offline refractivity grid	ORGMEC	"	"
GRM-75	Offline temperature grid	OTGMEC	"	"
GRM-76	Offline specific humidity grid	OHGMEC	"	"
GRM-77	Offline dry geopotential height grid	OZGMEC	"	"
GRM-78	Offline dry temperature grid	ODGMEC	"	"
GRM-79	Offline dry pressure grid	OYGMEC	"	"
GRM-193	Offline tropopause height grid	OCGMEC	"	"
GRM-83	Offline bending angle grid	OBGMET	Metop	netCDF
GRM-84	Offline refractivity grid	ORGMET	"	"
GRM-85	Offline temperature grid	OTGMET	"	"
GRM-86	Offline specific humidity grid	OHGMET	"	"
GRM-87	Offline dry geopotential height grid	OZGMET	"	"
GRM-88	Offline dry temperature grid	ODGMET	"	"
GRM-89	Offline dry pressure grid	OYGMET	"	"
GRM-194	Offline tropopause height grid	OCGMET	"	"

Table 1.4. List of *NTC data products* covered by this ATBD. The Level 1B input data to the ROM SAF processing is obtained from the EUMETSAT Secretariat.

Product ID	Product Name	Product acronym	Satellite input	File format
GRM-123	NTC bending angle grid	OBGS6	Sentinel-6	netCDF
GRM-124	NTC refractivity grid	ORGS6	"	"
GRM-125	NTC temperature grid	OTGS6	"	"
GRM-126	NTC specific humidity grid	OHGS6	"	"
GRM-127	NTC dry geopotential height grid	OZGS6	"	"
GRM-128	NTC dry temperature grid	ODGS6	"	"
GRM-129	NTC dry pressure grid	OYGS6	"	"
GRM-195	NTC tropopause height grid	OCGS6	"	"

1.2 Applicable and reference documents

1.2.1 Applicable documents

The following list contains documents with a direct bearing on the contents of this document:

- [AD.1] CDOP-3 Proposal: Proposal for the Third Continuous Development and Operations Phase (CDOP-3); Ref: SAF/ROM/DMI/MGT/CDOP3/001 Version 1.2 of 31 March 2016, Ref: EUM/C/85/16/DOC/15, approved by the EUMETSAT Council at its 85th meeting on 28-29 June 2016
- [AD.2] CDOP-3 Cooperation Agreement: Agreement between EUMETSAT and DMI on the Third Continuous Development and Operations Phase (CDOP-3) of the Radio Occultation Meteorology Satellite Applications Facility (ROM SAF), Ref. EUM/C/85/16/DOC/19, approved by the EUMETSAT Council and signed at its 86th meeting on 7 December 2016
- [AD.3] ROM SAF Product Requirements Document, SAF/ROM/DMI/MGT/PRD/001

1.2.2 Reference documents

The following documents provide supplementary or background information, and could be helpful in conjunction with this document:

- [RD.1] ROM SAF ATBD: Level 1B bending angles, SAF/ROM/DMI/ALG/BA/001.
- [RD.2] ROM SAF ATBD: Level 2A refractivity profiles, SAF/ROM/DMI/ALG/REF/001.
- [RD.3] ROM SAF ATBD: Level 2A dry temperature profiles, SAF/ROM/DMI/ALG/TDRY/001.
- [RD.4] ROM SAF ATBD: Level 2B and 2C 1D-Var products, SAF/ROM/DMI/ALG/1DV/002.
- [RD.5] ROM SAF ATBD: Level 2C tropopause height, SAF/ROM/DMI/ALG/TPH/001.
- [RD.6] ROM SAF Validation Report: Level 1B bending angle, Level 2A refractivity, Level 2A dry temperature, SAF/ROM/DMI/REP/ATM/001.
- [RD.7] ROM SAF Validation Report: Level 2B and 2C 1D-Var products, SAF/ROM/DMI/REP/1DV/002.
- [RD.8] ROM SAF Validation Report: Level 2C tropopause height, SAF/ROM/DMI/REP/TPH/001.
- [RD.9] The ROPP User Guide – Part III: Pre-processor module, SAF/ROM/METO/UG/ROPP/004.
- [RD.10] Lemoine, F. G., et al, The development of the joint NASA GSFC and NIMA geopotential model EGM96, NASA/TP-1998-206861, NASA GSFC, Greenbelt, MD, USA, 1998.
- [RD.11] Implications of using the World Geodetic System 1984 (WGS84), ET-SAT-6/Doc.16(1), WMO, Geneva, Switzerland, 12-15 April, 2011.
- [RD.12] Gleisner, H., Latitudinal binning and area-weighted averaging of irregularly distributed radio occultation data, GRAS SAF Report 10, 2010.

- [RD.13] Gorbunov, M.E., Ionospheric correction and statistical optimization of radio occultation data, *Radio Science*, **37**(5), 1084, 2002.
- [RD.14] Healy, S. and Eyre, J. R., Retrieving temperature, water vapor and surface pressure information from refractive-index profiles derived by radio occultation: A simulation study, *Quart. J. Roy. Meteorol. Soc.*, 126, 1661–1683, 2000.
- [RD.15] Healy, S., Refractivity coefficients used in the assimilation of GPS radio occultation measurements, GRAS SAF Report 09, 2009.
- [RD.16] Goody, R., Anderson, J., and North, G.: Testing climate models: An approach, *Bull. Amer. Meteor. Soc.*, 79(11), 2541-2549, 1998.
- [RD.17] Rieder, M. J., and Kirchengast, G., Error analysis and characterization of atmospheric profiles retrieved from GNSS occultation data, *J. Geophys. Res.*, vol. 106, 31755, 2001.
- [RD.18] Marquardt, C., Y. Andres, A. von Engeln, and F. Sancho (2009): SNR and bending angle noise in CHAMP, GRACE, COSMIC, and GRAS data, presentation at the COSMIC Workshop, Oct 2009, Boulder, USA.
- [RD.19] Foelsche, U., et al., Observing upper troposphere–lower stratosphere climate with radio occultation data from the CHAMP satellite, *Clim. Dyn.*, vol. 31, 49-65, 2008.
- [RD.20] Schreiner, W., S. Sokolovskiy, D. Hunt, C. Rocken, and Y.-H. Kuo, Analysis of GPS radio occultation data from the FORMOSAT-3/COSMIC and the Metop/GRAS missions at CDAAC, *Atmos. Meas. Tech.*, 4, 2255-2272, 2011.
- [RD.21] Scherllin-Pirscher, B., G. Kirchengast, A. K. Steiner, Y.-H. Kuo, and U. Foelsche, Quantifying uncertainty in climatological fields from GPS radio occultation: an empirical-analytical error model, *Atmos. Meas. Tech.*, 4, 2019-2034, 2011.
- [RD.22] Scherllin-Pirscher, B., A. K. Steiner, G. Kirchengast, Y.-H. Kuo, and U. Foelsche, Empirical analysis and modeling of errors of atmospheric profiles from GPS radio occultation, *Atmos. Meas. Tech.*, 4, 1875-1890, 2011
- [RD.23] Scherllin-Pirscher, B., S. Syndergaard, U. Foelsche, and K. B. Lauritsen, Generation of a bending angle radio occultation climatology (BAROCLIM) and its use in radio occultation retrievals, *Atmos. Meas. Tech.*, 8, 109-124, 2015.
- [RD.24] Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy, Observing Earth’s atmosphere with radio occultation measurements using the Global Positioning System, *J. Geophys. Res.*, 102, 23429-23465, 1997.
- [RD.25] Kuo, Y.-H., T.-K. Wee, S. Sokolovskiy, C. Rocken, W. Schreiner, D. Hunt, and R.A. Anthes, Inversion and error estimation of GPS radio occultation data, *J. Meteorol. Soc. Japan*, 82, 507-531, 2004.
- [RD.26] Smith, E.K., and S. Weintraub, The constants in the equation for atmospheric refractive index at radio frequencies, *Proc. IRE*, 41, 1035-1037, 1953.
- [RD.27] The ROPP User Guide – Part II: Forward module and 1D-Var modules, SAF/ROM/METO/UG/ROPP/003
- [RD.28] Ho, S.-P., et al., Estimating the uncertainty of using GPS radio occultation data for climate monitoring: Inter-comparison of CHAMP refractivity climate

records 2002-2006 from different data centers, *J. Geophys. Res.*, 114, D23107, 2009.

- [RD.29] Steiner, A. K., et al., Quantification of structural uncertainty in climate data records from GPS radio occultation, *Atmos. Chem. Phys.*, 13, 1469-1484, 2013.
- [RD.30] Desroziers, G., L. Berre, B. Chapnik, and P. Poli, Diagnosis of observation, background and analysis error statistics in observation space, *Q. J. Royal Meteorol. Soc.*, 131, 3385–3396, 2005

1.3 Acronyms and abbreviations

ATBD	Algorithm Theoretical Baseline Document
BAROCLIM	Bending Angle Radio Occultation CLIMatology
CHAMP	CHallenging Mini-satellite Payload (Germany)
CDAAC	COSMIC Data Analysis and Archive Center
CDOP-2	Second Continuous Development and Operations Phase (EUMETSAT)
CDR	Climate Data Record
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate (USA/Taiwan)
DMI	Danish Meteorological Institute; ROM SAF Leading Entity
ECMWF	The European Centre for Medium-range Weather Forecasts
EGM96	Earth Geopotential Model 96
EPS	EUMETSAT Polar System
EPS-SG	EUMETSAT Polar System – Second Generation
EUMETSAT	EUropean organisation for the exploitation of METeorological SATellites
GNSS	Global Navigation Satellite System
GPS	Global Positioning System (USA)
GRACE	Gravity Recovery and Climate Experiment
GRAS	GNSS Receiver for Atmospheric Sounding (EPS/Metop)
ICDR	Interim Climate Data Record
IIEC	Institut d’Estudis Espacials de Catalunya
LEO	Low-Earth Orbit
Met Office	United Kingdom Meteorological Office
Metop	Meteorological Operational Polar satellite (EUMETSAT)
netCDF	network Common Data Form (Unidata)
NRT	Near Real Time
NTC	Non Time Critical
PUM	Product User Manual
RO	Radio Occultation
ROM SAF	Radio Occultation Meteorology SAF (former GRAS SAF)
ROPP	Radio Occultation Processing Package
SAF	Satellite Application Facility (EUMETSAT)
UCAR	University Corporation for Atmospheric Research (Boulder, CO, USA)
VAR	Variational (NWP data assimilation technique)
WGS-84	World Geodetic System 1984
WMO	World Meteorological Organization
WWW	World Weather Watch (WMO programme)

1.4 Definitions

RO data products from the Metop, Metop-SG and Sentinel-6 satellites and RO data from other missions are grouped in *data levels* (level 0, 1, 2, or 3) and *product types* (NRT, Offline, NTC, CDR, or ICDR). The data levels for atmospheric data¹ and product types are defined below². The lists of variables should not be considered as the complete contents of a given data level, and not all data may be contained in a given data level.

Data levels:

Level 0: Raw sounding, tracking and ancillary data, and other GNSS data before clock correction and reconstruction;

Level 1A: Reconstructed full resolution excess phases, total phases, pseudo ranges, SNRs, orbit information, I, Q values, NCO (carrier) phases, navigation bits, and quality information;

Level 1B: Bending angles and impact parameters, tangent point location, and quality information;

Level 2: Refractivity, geopotential height, “dry” temperature profiles (Level 2A), pressure, temperature, specific humidity profiles (Level 2B), surface pressure, tropopause height, planetary boundary layer height (Level 2C), ECMWF model level coefficients (Level 2D), quality information;

Level 3: Gridded or resampled data that are processed from Level 1 or 2 data, and that are provided as, e.g., daily, monthly, or seasonal means on a spatiotemporal grid, including metadata, uncertainties and quality information.

Product types:

NRT product: Data product delivered less than: (i) 3 hours after measurement (ROM SAF Level 2 for EPS); (ii) 150 min after measurement (ROM SAF Level 2 for EPS-SG Global Mission); (iii) 125 min after measurement (ROM SAF Level 2 for EPS-SG Regional Mission);

Offline and NTC products: Data product delivered from about 5 days to up to 6 months after measurement, depending on the applicable requirements. The evolution of this type of product is driven by new scientific developments and subsequent product upgrades;

CDR: Climate Data Record generated from a dedicated reprocessing activity using a fixed set of processing software³. The data record covers an extended time period of several years (with a fixed end point) and constitutes a homogeneous data record appropriate for climate usage;

¹ Definitions for ionosphere products will be included when preparing for the RR review.

² Note that the level definitions differ partly from the WMO definitions:

http://www.wmo.int/pages/prog/sat/dataandproducts_en.php

³ (i) GCOS 2016 Implementation Plan; (ii) <http://climatemonitoring.info/home/terminology/>

ICDR: An Interim Climate Data Record (ICDR) regularly extends in time a (Fundamental or Thematic) CDR using a system having optimum consistency with and lower latency than the system used to generate the CDR⁴.

General terms:

System: GPAC (ROM SAF GNSS Processing and Archiving Center)

Web site: ROM SAF web site: <http://www.romsaf.org>

Product Archive: PARF (ROM SAF Product Archive and Retrieval Facility)

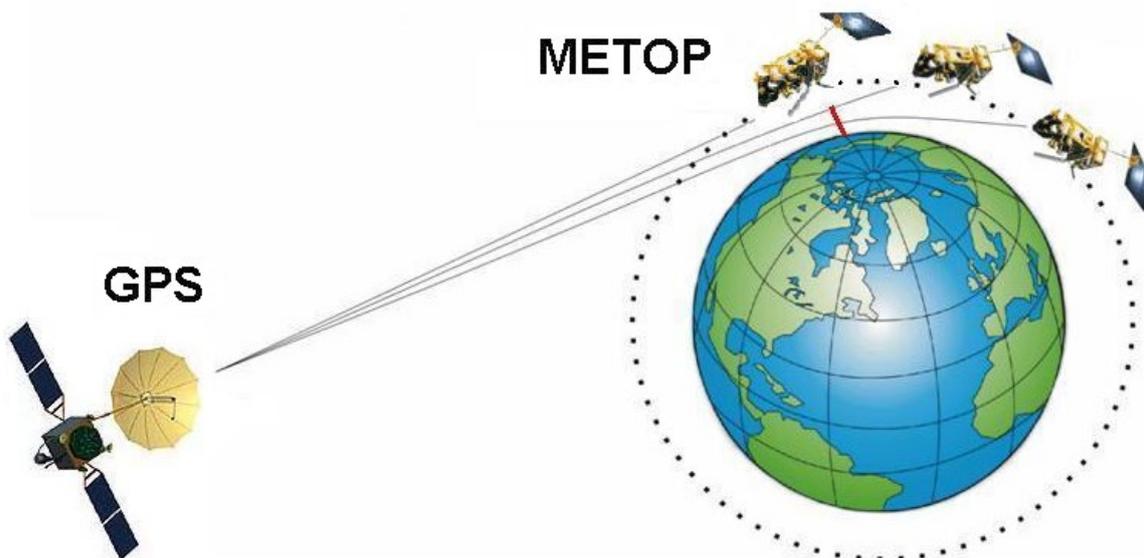
⁴ <http://climatemonitoring.info/home/terminology/> (the ICDR definition was endorsed at the [9th session of the joint CEOS/CGMS Working Group Climate Meeting on 29 March 2018](#))

2. The radio occultation technique

A radio occultation (RO) instrument on board a low-Earth orbit (LEO) satellite, e.g. the GRAS instrument on board a Metop satellite, measures the phase and amplitude of the Doppler shifted radio signals from a GNSS satellite as the satellite sets or rises behind the Earth's limb (Figure 1). Through a rather complex sequence of processing steps the Doppler shift and amplitude as a function of time is converted to a neutral-atmosphere profile of bending angle, which is inverted to refractivity, dry pressure and dry temperature, and finally to physical pressure, temperature, and humidity [RD.1-5]. The latter three quantities are obtained through a 1D-Var procedure which requires *a priori* information taken from an atmospheric model [RD.4]. *A priori* data are also used for bending angle initialization in order to retrieve refractivity [RD.1].

The Level 3 gridded RO data products are based on the ROM SAF Level 1B and 2 data products. The aim of the Level 3 processing algorithms is to produce spatial and temporal averages – currently zonal monthly means on a 5 degree by 200 meter latitude-height grid – as well as other statistics of the observed RO profile data, and to describe the uncertainties in the averages. The algorithms are designed to obtain true area-weighted means from a set of irregularly distributed observations of the atmosphere, and to address the limitations due to under-sampling. Temporal under-sampling is a concern that is common for most observations made from LEO orbit – it has to do with the particular vantage point in space, rather than any instrument characteristics – whereas the methods used to achieve a correct area-weighting depends on how the particular instrument scans the Earth's atmosphere.

The methods described in this report have been used to generate gridded monthly-mean RO data from the CHAMP, GRACE, COSMIC, Metop, and Sentinel-6 missions. The examples in this document are taken from these data.



3. Height variables

The *World Geodetic System 1984 (WGS-84)* comprises a *terrestrial reference frame* and an associated *reference ellipsoid*. It is the reference system used by the GPS system. To connect observations made using GPS signals with other type of observations and with geophysical model data, a *geoid* needs to be defined in WGS-84 coordinates. The most commonly used geoid model within the RO community is based on the Earth gravitational field model EGM-96 [RD.10]. This is also the geoid model recommended by WMO to be used as a fixed reference for mean-sea level (MSL) determinations [RD.11].

The relationship between ellipsoidal height, h , and MSL altitude, H , (alternatively referred to as orthometric height) is

$$H = h - u$$

where u is the *geoid undulation*, i.e. the ellipsoidal height of the geoid. As shown in Figure 2, h is the straight-line distance, whereas H is measured along the slightly curved plumb line. Hence, Eq. 1 is an approximation, but since the angle ε between the ellipsoidal normal and the plumb line is very small (from a few arc-seconds to a few tens of arc-seconds near the Earth's surface), the approximation is sufficiently accurate for all practical purposes.

The geoid is per definition the surface where the gravitational potential, Φ , is everywhere zero. The gravitational potential at a point in space above the geoid is given by the integral of the gravitational acceleration, $g(H)$, along a plumb line. Hence, the Earth's gravitational potential, or geopotential, can effectively be used to express height above the geoid. We define the geopotential height, z , as

$$z(H) \equiv \frac{\Phi(H)}{g_{\text{WMO}}}$$

where g_{WMO} is assigned a constant value 9.80665 m/s^2 according to a WMO convention. Note that even though geopotential height has the unit meter, it is not a measure of height in a strict sense but rather a measure of the geopotential.

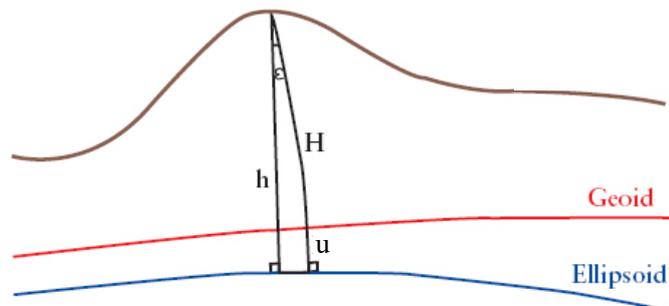


Figure 2. Relations between ellipsoidal height (h), MSL altitude (H) and geoid undulation (u). While h is a straight-line distance, H is measured along the slightly curved plumb line. The vertical deflection angle ε is small, and for all practical purposes H can be taken as a straight-line distance.

The inter-relations between h , H , and z are thus determined by a certain choice of gravitational model, i.e. a model for the geoid in combination with a model for the geopotential, or the gravitational acceleration. In the ROM SAF Level 3 processing, the geoid undulation is computed from the NASA/NIMA EGM96 spectral model with respect to the WGS84 ellipsoid. This model is provided in the form of geoid potential and correction coefficients to order and degree 360 [RD.10]. The coefficients are expanded as Legendre polynomials and applied to the reference location of the occultation.

Atmospheric pressure is a monotonic function of geopotential height. Alternatively, the geopotential height can be described as a function of pressure (or dry pressure, which is explained in Sections 4.1 and 4.2), in which case we denote it by capital Z (or Z_{dry} for geopotential height of dry pressure). If the atmospheric pressure is used as the independent variable, it is often practical to transform it to a *pressure height* [RD.19], here defined as

$$H_p = H_{\text{atm}} \cdot \ln\left(\frac{p_{\text{atm}}}{p}\right) \quad (3)$$

where p_{atm} is a constant set to 1013.25 hPa and H_{atm} is a constant atmospheric scale height set to 7000 meter (see Table 3.1). Note that even though the pressure height has the unit meter, it is a measure of atmospheric pressure rather than a measure of height.

<i>Pressure height</i>	<i>Pressure</i>
0 km	1013.25 hPa
5 km	496.03 hPa
10 km	242.83 hPa
15 km	118.87 hPa
20 km	58.19 hPa
30 km	13.95 hPa
40 km	3.34 hPa

Table 3.1. The atmospheric pressure at different pressure heights, according to Eq. 3.

4. Level 1B and Level 2 profile data

4.1 From measurements to atmospheric profiles

The fundamental observable measured by an RO instrument is the phase, L_i , and amplitude, Φ_i , of the Doppler-shifted incoming signal. Index i denotes a GNSS frequency, currently L1 and either L2 or L5. It is outside the scope of the present document to describe the processing from measured phase and amplitude to atmospheric profiles in any detail. Comprehensive descriptions can be found in [RD.1,2,9]. The ROPP User Guide Part III [RD.9] gives a detailed account of the algorithms used in the ROM SAF processing. In the present document, we only draw a rough sketch and particularly point out in which steps *a priori* data, taken from an atmospheric model, are allowed to influence the observed data.

From phase and amplitude measurements, and the satellites' positions and velocities, we obtain the observed bending angles as a function of impact parameter at the GNSS frequencies L1 and L2/L5 [RD.1,9] (see Section 4.2 for a discussion of the impact parameter and related concepts). The influence of the ionosphere can be removed to first order by forming a linear combination of the bending angles at two frequencies thus obtaining the observed (LC) bending angle α_{obs} . This observed bending angle is contaminated with noise that increases exponentially with altitude rendering it useless above a certain height. However, we need bending angles to infinite altitudes in order to obtain the refractivity. The solution is to form a statistically optimal linear combination of the observed bending angle, α_{obs} , and a background bending angle, α_{bg} , according to

$$\alpha(a) = \alpha_{\text{bg}}(a) + \frac{\sigma_{\text{bg}}^2}{\sigma_{\text{bg}}^2 + \sigma_{\text{obs}}^2} (\alpha_{\text{obs}}(a) - \alpha_{\text{bg}}(a)) \quad (4)$$

where the relative contributions of observation and background – or *a priori* data – are determined by the errors σ_{bg} and σ_{obs} (the latter is an estimate of the error in the observed LC bending angle). The error models are chosen such that the fraction goes from no background at low altitudes to no observational information at high altitudes

The *a priori* data in Eq. 4 is taken from the BAROCLIM climatological model, which represents the monthly-mean atmospheric state from 2006 to 2012 as measured by the FORMOSAT-3/COSMIC mission [RD.23]. Currently, the background profile in Eq. 4 is obtained by searching BAROCLIM for a profile that fits a smoothed version of the observed bending angle in the interval 40 to 60 km. Before merging the observed and background bending angles, the logarithm of the background profile is scaled and shifted to avoid introducing biases into the statistically optimized bending angle. In the ROM SAF processing, the ionospheric correction and the statistical optimization steps are combined into a single framework that takes a somewhat more complex form than in the above equation [RD.1,11].

Bending angle is related to the vertical gradient of the refractive index n . The relation can be inverted using the inverse Abel transform to give the refractive index as a function of height from knowledge of the bending angles as a function of impact parameter [RD.1,9]. The refractive index is expressed in terms of the refractivity, defined as

$$N \equiv (n - 1) \cdot 10^6$$

which can be regarded as an ordinary physical state variable since it is a function of other state variables. The expression used by the RO community is commonly

$$N = \frac{\kappa_1}{Z_d} \frac{p_d}{T} + \frac{\kappa_2}{Z_w} \frac{p_w}{T^2} + \frac{\kappa_3}{Z_w} \frac{p_w}{T} \quad (6)$$

where p_d and p_w are the partial pressures of dry air and water vapour, respectively, and Z_d and Z_w are the corresponding non-ideal compressibility factors, which for ideal gas are unity. The standard values for the coefficients in Eq. 6 are $\kappa_1=77.6$ K/hPa, $\kappa_2=3.73 \cdot 10^5$ K²/hPa, and $\kappa_3=77.6$ K/hPa [RD.26]. Recent investigations have indicated the need for a small revision of these values, and have shown a weak dependence on deviations from the ideal gas law [RD.15]. When using a non-ideal gas law in applications that include forward modelling from model data, the coefficients used are instead $\kappa_1=77.643$ K/hPa, $\kappa_2=3.75463 \cdot 10^5$ K²/hPa, and $\kappa_3=71.2952$ K/hPa, and the compressibility factors may be up to 0.05% smaller than unity in the denser parts of the atmosphere [RD.27].

When humidity is negligible, the second and third terms on the right hand side in Eq. 6 vanish, and for an ideal gas the refractivity is directly proportional to the air density. Using the equation of state for an ideal gas and assuming hydrostatic equilibrium, the dry pressure profile is obtained by integrating a version of the hydrostatic equation

$$\frac{d \ln p}{dH} = - \frac{g(H)N(H)}{R\kappa_1 p(H)} \quad (7)$$

from an upper boundary where the pressure is assumed to be known. Here, H is the mean-sea level altitude, $g(H)$ is the gravitational acceleration, and R is the gas constant for dry air. Dry temperature is computed from the dry pressure and the observed refractivity (using Eq. 6 with the “wet” term ignored). The pressure prescribed at the upper boundary is given by

$$p(H_{\text{top}}) = - \frac{g(H_{\text{top}})N(H_{\text{top}})}{R\kappa_1 \cdot d \ln N/dz} \quad (8)$$

which assumes that $dT/dz = 0$ at H_{top} , taken to be 150 km. If the refractivity profile does not reach 150 km, it is exponentially extrapolated to that height.

The $\ln(N)$ profile is first interpolated (using quadratic spline interpolation) to a high-resolution integration grid with a 15 m step size. The dry pressure profile is obtained by a fourth order

Runge-Kutta integration scheme. The dry temperature profile is computed from N and p , and T and $\ln(p)$ are then interpolated back to the original levels with an approximate 100 m spacing.

In the troposphere the influence from water vapour on the observed refractivity is not negligible. We thus have a temperature-humidity ambiguity which can only be resolved by introducing additional data on temperature and humidity. This is done through a 1D variational (1D-Var) procedure in which the observed refractivity profile is combined with a model profile in a statistically optimal way considering the errors and vertical error correlations of both the observations and the *a priori* data [RD.4,14]. A solution is found by minimizing the cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - \mathbf{H}(\mathbf{x}))^T \mathbf{O}^{-1}(\mathbf{y}^o - \mathbf{H}(\mathbf{x})) \quad (9)$$

with respect to the atmospheric state \mathbf{x} . \mathbf{H} is the forward operator mapping the atmospheric state \mathbf{x} into measurement space, and \mathbf{O} and \mathbf{B} are the observation and background error covariance matrixes, respectively. The *a priori* data \mathbf{x}^b used in the ROM SAF 1D-Var processing are interpolated from ECMWF reanalysis short-term forecast fields at a $1.0^\circ \times 1.0^\circ$ degree horizontal resolution. The reanalysis used is either ERA-Interim, for CDR v1.0, or for later data versions it is a combination of ERA5.1 before 2007 and ERA5 from 2007 and onwards

For each occultation, the tropopause height is computed from the profile data. The ROM SAF tropopause height product is based on the dry temperature lapse rate according to a WMO definition of tropopause [RD.5].

<i>a priori</i>	influencing Level 3 data product
<i>None</i>	Bending angle
BAROCLIM climatology	Refractivity
BAROCLIM climatology	Dry temperature
BAROCLIM climatology	Dry pressure
BAROCLIM climatology	Dry geopotential height
ECMWF reanalysis forecasts	1D-Var temperature
ECMWF reanalysis forecasts	1D-Var humidity

Table 4.1. A priori data influence the Level 3 gridded data products through the statistical optimization of the bending angles and through the 1D-Var procedure described in Section 4.1. The ECMWF reanalysis is either ERA-Interim, for CDR v1.0, or for later data versions it is a combination of ERA5.1 before 2007 and ERA5 from 2007 and onwards. The Level 3 bending angles are based on the observed (i.e., “raw”, not statistically optimized) bending angles.

4.2 Atmospheric profile data

The Level 1B and Level 2 profile data consist of:

- Bending angle, $\alpha(a)$
- Refractivity, $N(H)$
- Dry pressure, $p_{\text{dry}}(H)$
- Dry temperature, $T_{\text{dry}}(H)$
- Dry geopotential height, $Z_{\text{dry}}(H_p)$
- 1D-Var temperature, $T(H)$
- 1D-Var specific humidity, $q(H)$

and from the dry-temperature profile, the tropopause height is derived:

- Tropopause height, H_{TP}

Here, H is the MSL altitude defined in Section 3, and a is the impact parameter. The bending angle and the microwave refractivity are somewhat unfamiliar physical variables for most atmospheric scientists outside the RO community. The bending angle is an integral property of a ray at radio frequencies, where each ray is referenced by the impact parameter, a . Under the assumption that the atmosphere is spherically symmetric around the occultation point, the impact parameter, a , can be related to the ray tangent height, r_t , and the atmosphere's refractive index, n , through

$$a = n \cdot r_t \quad (10)$$

The tangent height, r_t , is measured with respect to the center of curvature of the spherically symmetric atmosphere and can be described in terms of the local radius of curvature, R_c , of the Earth ellipsoid and the geometric height, h , with respect to the Earth ellipsoid

$$r_t = R_c + h = R_c + H + u \quad (11)$$

where H is the MSL altitude, and u is the geoid undulation.

The impact altitude, H_a , is used as an alternative to the impact parameter. It is defined as

$$H_a \equiv a - R_c - u$$

and differs from the mean-sea level altitude, H , by

$$H_a - H = (n - 1) \cdot r_t \quad (13)$$

which amounts to a few meters in the upper stratosphere, where the refractive index is near unity. Around the tropopause, H and H_a differ by around 500 meters and deep down in the troposphere, near the surface, they differ by up to 1 to 2 kilometers. The impact altitude defined by Eq. 12 differs from the commonly used impact height, H_i , only through the subtraction of the geoid undulation u , i.e. $H_i = a - R_c$.

The atmospheric profiles listed above are described as functions of MSL altitude. Alternatively, the profiles can be described as functions of geopotential height z . Since the atmospheric pressure, $p(z)$, is a monotonic function of height, we can invert the functional relation to $z(p)$. Hence, the geopotential height can be described as a function of pressure rather than the other way round. In many applications, it is common practice to describe the pressure field in terms of the geopotential heights of isobaric surfaces. In the current report, Z (with a capital letter) indicates that geopotential height is the independent variable.

Figure 3 shows a few examples of profiles measured by COSMIC in January 2009.

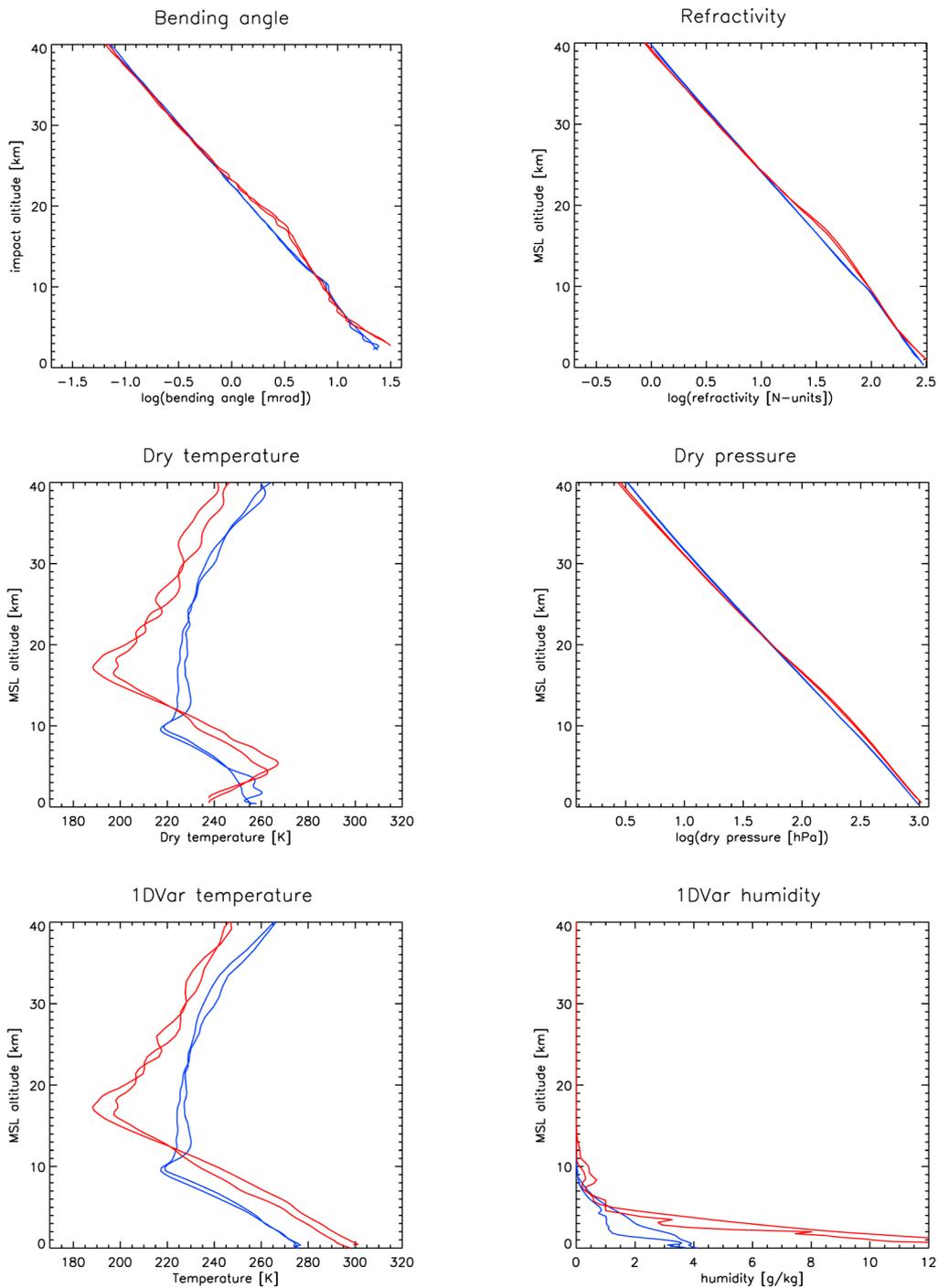


Figure 3. Vertical profiles of bending angle, refractivity, dry temperature, dry pressure, 1D-Var temperature, and 1D-Var humidity for 4 occultations observed by COSMIC in January 2009. The red profiles are from low latitudes, while blue profiles are from mid-latitudes.

4.3 Time and location of atmospheric profiles

A setting occultation normally takes 1 to 3 minutes, and in some cases even longer, from well above the neutral atmosphere until signal is lost near the Earth's surface. This is also the time during which a rising occultation is tracked from the moment the signal is detected at low altitudes. The neutral-atmosphere profile obtained is not a straight, vertical line but rather a 3-dimensional, slightly bent curve, which means that the actual latitude and longitude of the profile are functions of height. However, for the Level 3 processing and other purposes each profile is labelled with a single reference time and location.

The reference time for a profile is chosen as the start of the occultation. Due to the short duration of an occultation, and the large number of occultation events in each monthly grid box, the exact definition of the reference time has no practical influence on the climate data.

The reference location for a profile is obtained by following the straight line between the transmitting GNSS satellite and the receiving RO satellite until it touches the surface of the ellipsoid. The point of first contact (or last contact for rising occultations) between the straight line and the ellipsoid defines the reference latitude and longitude of the profile. This reference location corresponds roughly to the actual location of the profile (defined as the latitude and longitude of the ray tangent point) at an altitude of 5 to 15 kilometer.

4.4 Uncertainty of atmospheric profiles

The *measurement error* associated with a single profile can be considered purely random. We can think of the measurement error as a stochastic variable with zero mean and standard deviation s_{meas} . This standard deviation is the *measurement uncertainty*. We also assume that the random measurement errors of different profiles are independent (no error correlations between occultations). When averaging over a number of profiles, the measurement uncertainty of the resulting average decreases with $N^{1/2}$, where N is the number of data points included in the average.

The dominating error sources, and the main characteristics, of the random measurement errors are well understood (e.g., Kursinski et al., 1997) [RD.24]. RO data have smallest relative errors from the upper troposphere to the middle stratosphere. Kursinski et al. reported a refractivity error of 0.2%, which was confirmed by Schreiner et al. (2011) [RD.20]. Kuo et al. (2004) reported a refractivity error of around 0.3-0.5%, and Scherllin-Pirscher et al. (2011) stated an error in the upper troposphere-middle stratosphere region of about 0.35% [RD.22,25]. This value increases downward to around 3% near the surface (e.g., Kuo et al., 2004) [RD.25]. The reported bending angle errors are a factor of 2.5-3 larger (e.g., Scherllin-Pirscher et al., 2011), in line with expectations of how errors propagate in the RO processing chain (Rieder and Kirchengast, 2001) [RD.17,22]. We also expect the dry properties to have roughly the same relative errors as the refractivity.

From a minimum value in the upper troposphere-lower stratosphere region, the measurement relative errors increase upward, as the absolute errors attain a constant value governed by instrumental noise and ionospheric residual errors. These constant upper-level errors can be evaluated by examining the noise in a suitable height interval, e.g., 60-80 kilometer. As described in Appendix I, this is done as a part of the ROM SAF validation procedures. Schreiner et al. (2011) showed that the constant noise level (random measurement error) is around 1.8 μrad and 1.1 μrad for the COSMIC and Metop missions, respectively [RD.20].

Based on these results from the scientific literature, together with previous validation of ROM SAF RO data, a set of analytical functions (of altitude) has been defined that describes the measurement uncertainties of bending angle, refractivity, and the dry variables. In the ROM SAF formulation, the uncertainty for a bending angle profile decreases linearly from 6% to 0.9% between impact altitudes 0 kilometers and 10 kilometers. Above that, the uncertainty is 0.9% or 1.5 μrad , whichever is greatest. The corresponding uncertainties in the refractivity profiles are from 2% near the surface to 0.3% at 10 km, and above that the greatest of 0.3% and 0.01 N-units. The dry variables are similarly defined by a given relative-error profile limited by a fixed absolute error (for quantities that decrease exponentially with altitude) or by an exponentially increasing absolute error (for quantities that do not have an exponential fall-off with altitude).

The measurement uncertainty of variables obtained by 1D-Var retrievals – temperature, pressure, and humidity – are the formal errors obtained as a part of the 1D-Var solution. This is described in detail in [RD.4].

The measurement uncertainties for the bending-angle, refractivity, and dry profiles are thus formulated based on a fixed relative-error profile

$$s^{\text{rel}} = 0.06 + (0.009 - 0.06) \cdot \min\{H/10.0, 1.0\} \quad (14)$$

where H is altitude (or, alternatively, impact altitude or pressure altitude) expressed in kilometer. The measurement uncertainties for bending angle (s_α), refractivity (s_N), dry temperature (s_{T_d}), dry pressure (s_{P_d}) and dry geopotential height (s_{Z_d}) are given by

$$s_\alpha = \max\{ \alpha s^{\text{rel}}, 1.5 \mu\text{rad} \} \quad (15)$$

$$s_N = \max\{ N s^{\text{rel}}/3, 0.01 \text{ N-units} \} \quad (16)$$

$$s_{T_d} = \max\{ T_d \cdot s^{\text{rel}}/3, 12 \cdot \exp[(H-50)/10] \text{ K} \} \quad (17)$$

$$s_{P_d} = \max\{ p_d \cdot s^{\text{rel}}/6, 0.05 \text{ hPa} \} \quad (18)$$

$$s_{Z_d} = \max\{ 6500 s^{\text{rel}}/6, 100 \cdot \exp[(H_p-50)/14] \text{ gpm} \} \quad (19)$$

The profile error description used in the Level 3 processing does not currently include vertical error correlations. Neither is there any explicit dependence on latitude or season. However, as evident in Figures 4a-d, there is an implicit variation with latitude and season due to the dependence of the uncertainties on the profiles themselves.

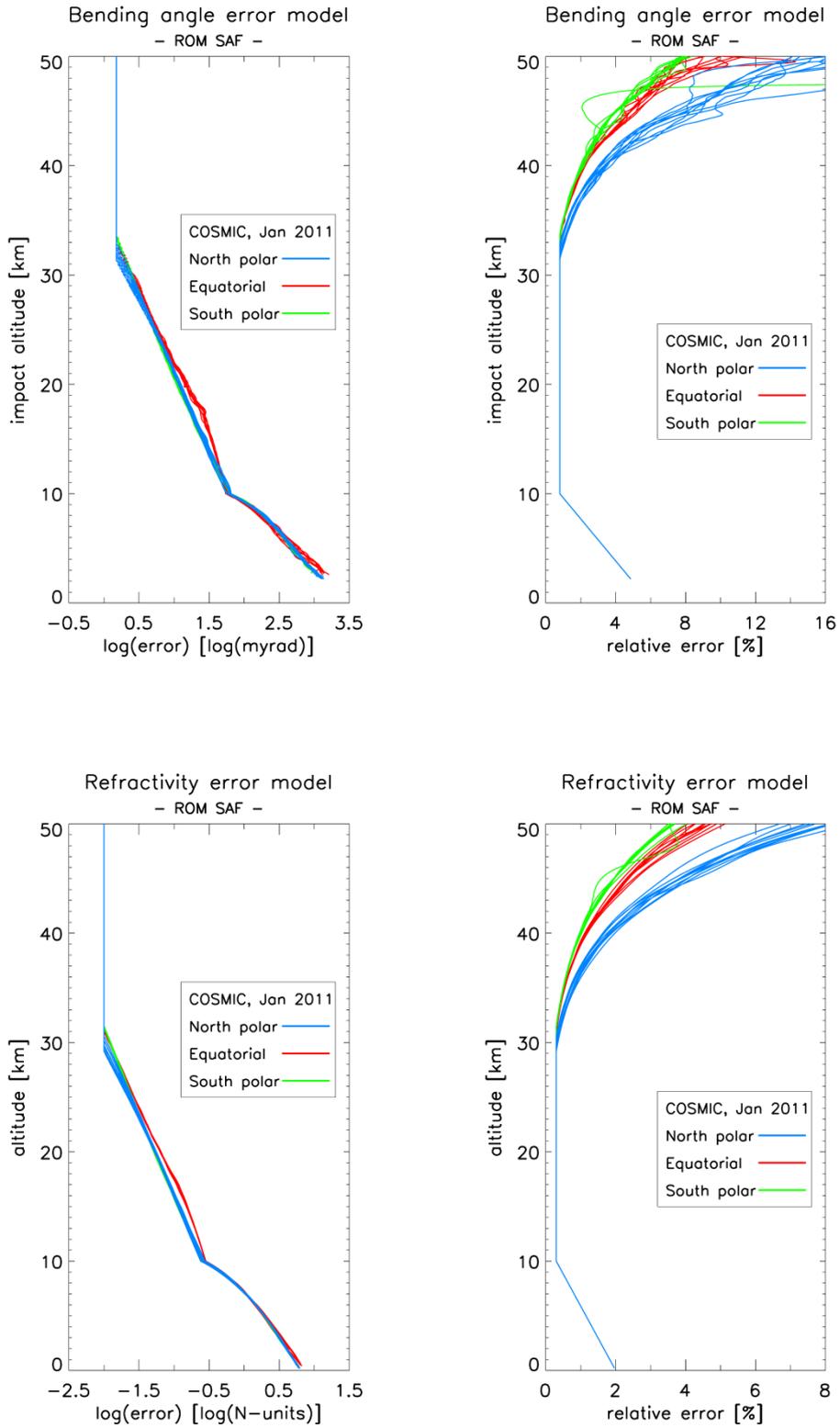


Figure 4a. Bending-angle and refractivity measurement uncertainties used in the ROM SAF Level 3 processing, here shown for a few COSMIC profiles from January 2011. The ROM SAF measurement error model is formulated as a given relative-error profile, but with a floor defined by a specified absolute error.

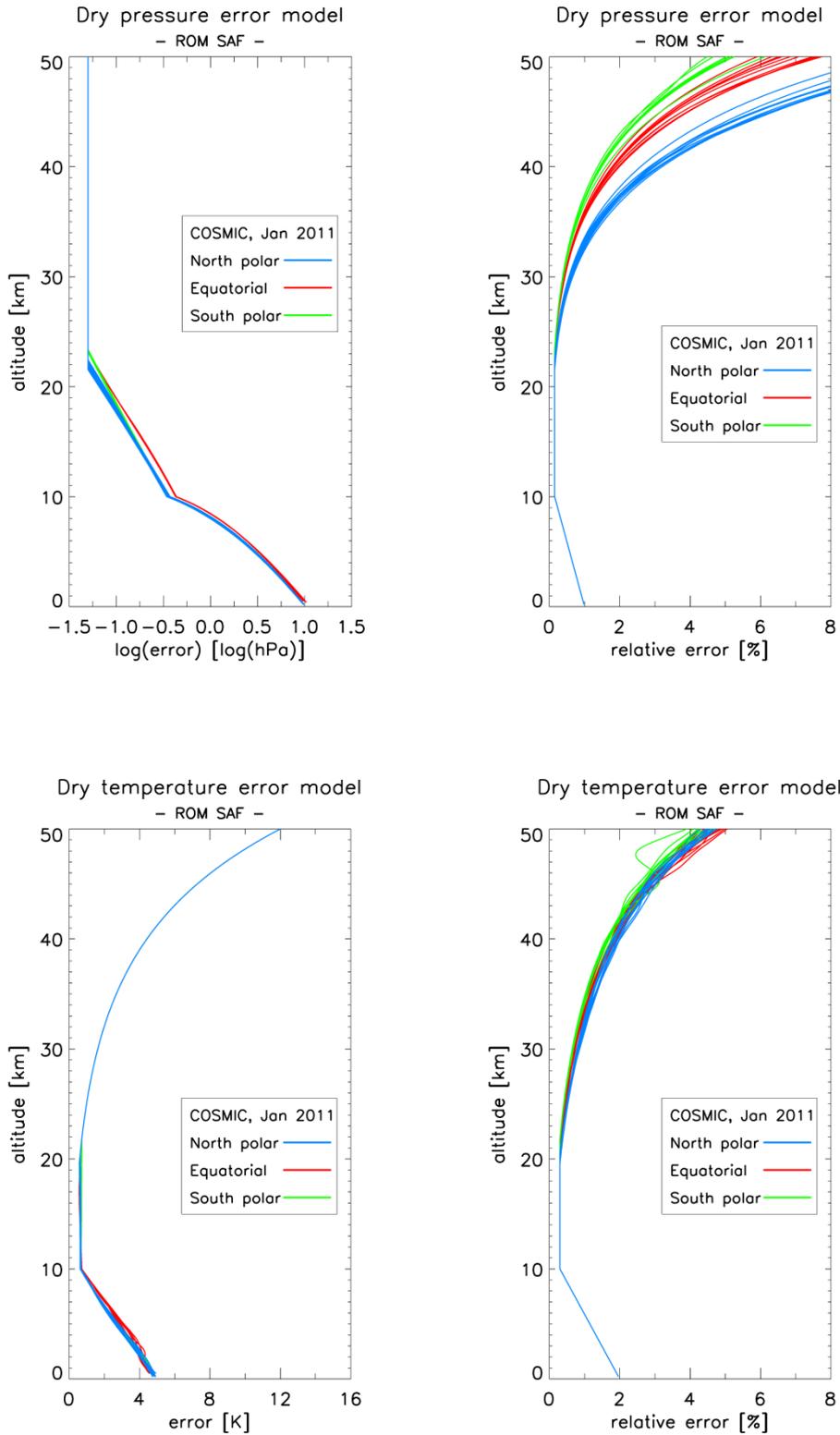


Figure 4b. Dry pressure and dry temperature measurement uncertainties used in the ROM SAF Level 3 processing, here shown for a few COSMIC profiles from January 2011. The ROM SAF measurement error model is formulated as a given relative-error profile, but with a floor defined by a specified absolute error.

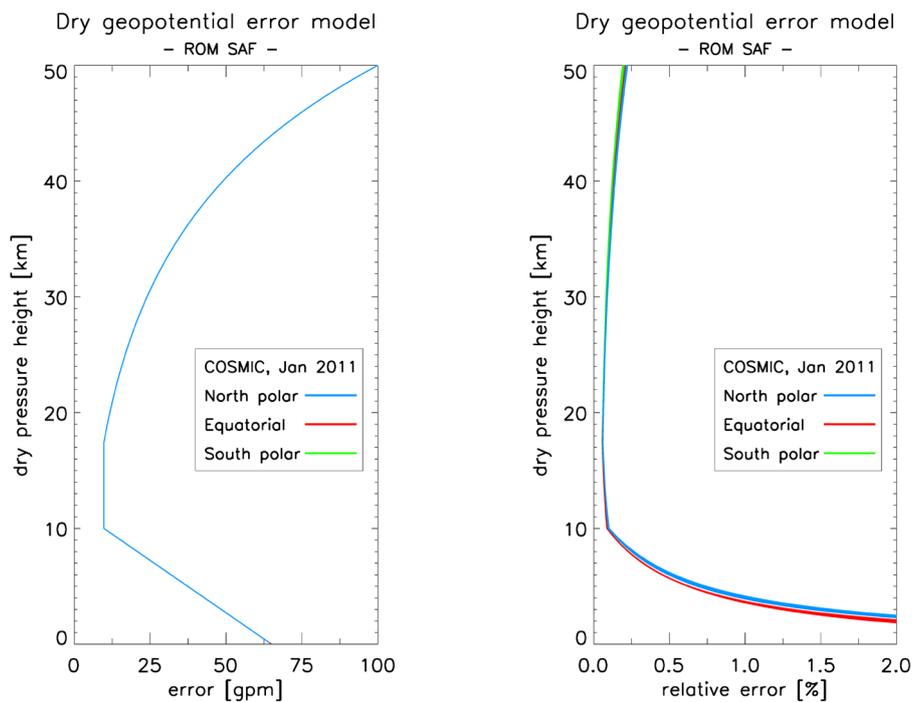


Figure 4c. Dry geopotential-height measurement uncertainties used in the ROM SAF Level 3 processing, here shown for a few COSMIC profiles from January 2011. The ROM SAF measurement error model is formulated as a given relative-error profile, but with a floor defined by a specified absolute error.

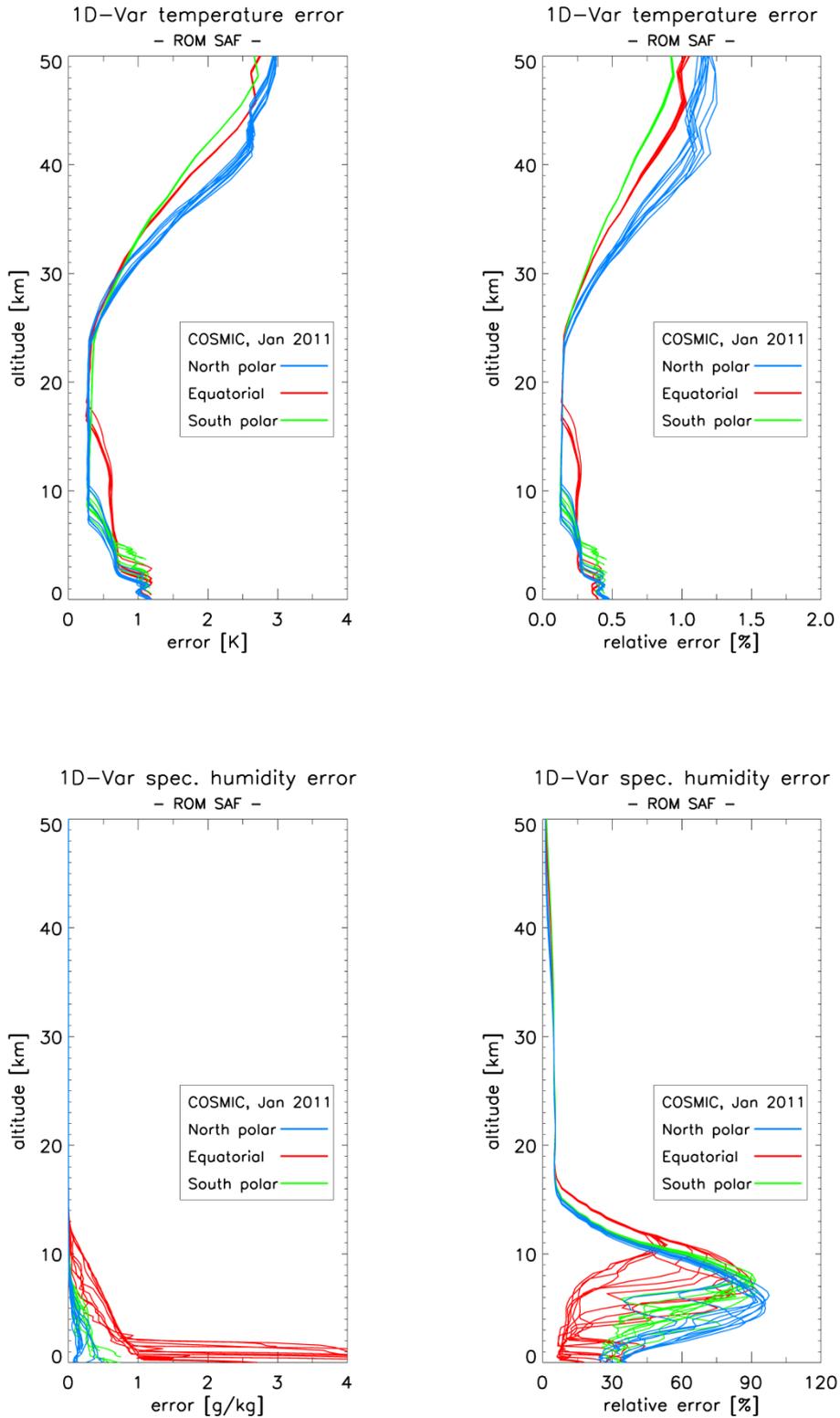


Figure 4d. Temperature and specific humidity measurement uncertainties from the 1D-Var retrievals used in the ROM SAF Level 3 processing, here shown for a few COSMIC profiles from January 2011. The measurement uncertainties are the combined result of the assumed observational uncertainties and the uncertainties in the background data.

4.5 Spatial and temporal sampling of the atmosphere

The exact time and locations of individual occultation events, as well as the statistical distribution of occultations, depend on the orbits of both the transmitting GNSS satellite and the LEO satellite which carries the RO receiver. As an example, the Metop satellites are in near-polar, low-Earth orbits with an orbital time of around 100 minutes. Observations made from such orbits have certain features in common: a near-global spatial coverage, approximately uniform in longitude but with a higher density toward the poles, and with characteristic spatio-temporal structures. The latter depends primarily on the nodal precession rate of the orbit which governs the temporal coverage of observations.

Figure 5 demonstrates the geographic location of occultation events during one day and during one month, for observations made by the GRAS instrument on board the Metop satellite. The tendency to a higher density of occultations at high latitudes, with correspondingly fewer events at low latitudes, is clearly seen.

In Figure 6, the number of observations per longitude and latitude bin is shown, together with the number of observations per unit area as a function of longitude and as a function of latitude. For the latitudinal distribution there is a marked difference between the two type of distributions. Detailed studies of the distribution of occultation events across 5-degree latitude grid boxes also show that in general, the latitudinal distribution across grid boxes is somewhat more uniform per area unit than per degree of latitude, particularly toward the poles [RD.12].

The Metop orbit has a precession rate that precisely matches the Earth's orbit around the Sun, keeping the orbital plane at a fixed angle to the Sun-Earth line (i.e. it is Sun-synchronous). This means that the local time

$$LT = UT + \lambda / 15 \quad (20)$$

for equatorial crossings does not change with universal time, UT . In Eq. 20, λ is the longitude expressed in degrees and UT is given in hours. Near the equator, the Metop observations are always made during a few hours around 9:30 in the morning and a few hours around 21:30 in the evening.

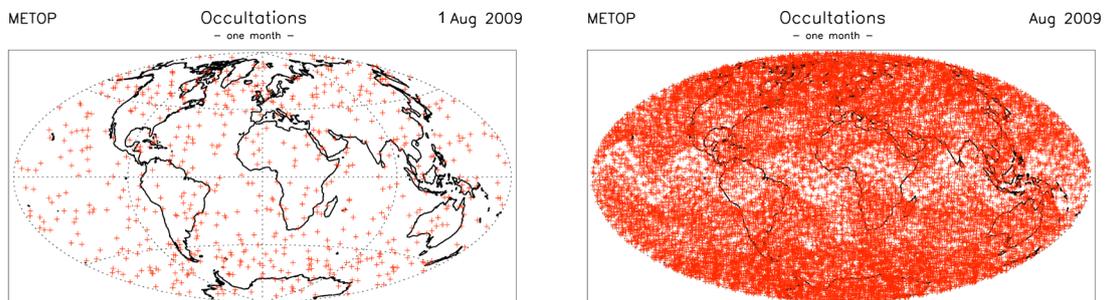


Figure 5. Geographic distribution of occultations observed by the GRAS instrument onboard Metop during one day (left panel: 604 occultations on August 1, 2009) and one month (right panel: 19369 occultations in August 2009). The map is an equal-area projection which means that apparent density variations correspond to actual variations in the number of events per unit area. Note the comparatively lower density at low latitudes.

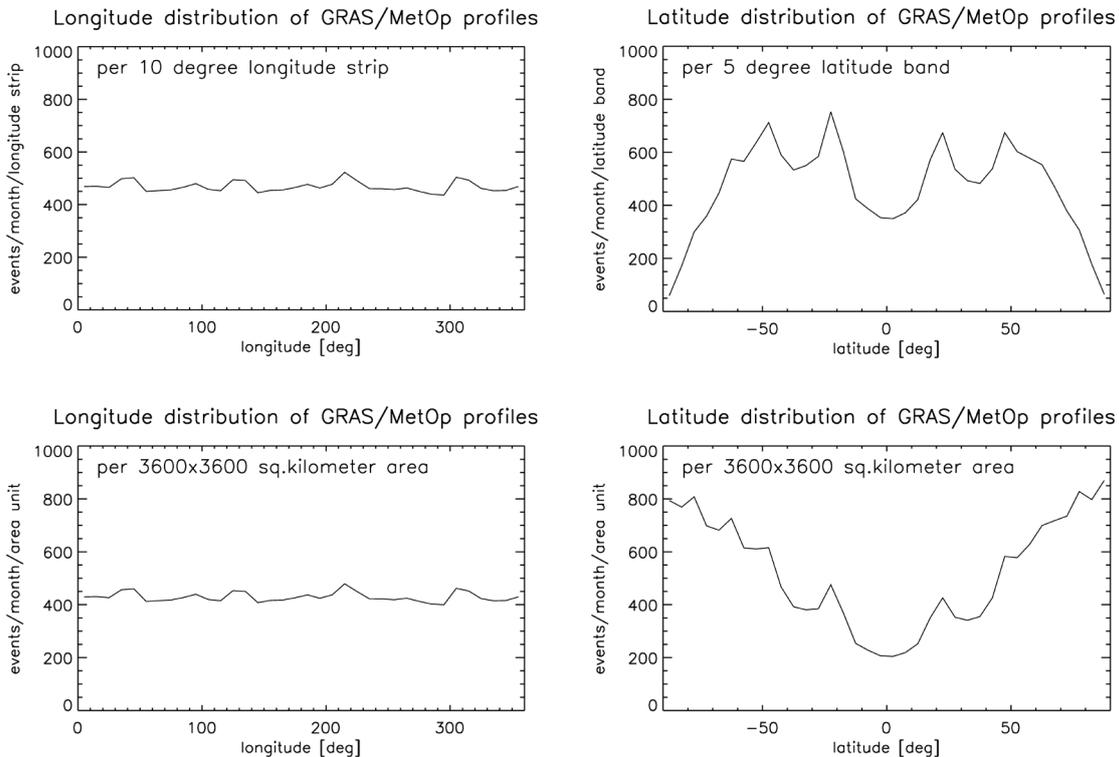


Figure 6. Spatial distributions of occultation events during the 3-month period July-September 2009. The upper panels show the number of occultation events per month distributed into 36 longitude bins and 36 latitude bins. The lower panels show the same distributions viewed per area unit instead of per equal-angle latitude band and longitude strip. In these panels, the longitude distributions are means over all latitudes and the latitude distributions are means over all longitudes.

Figure 7 shows that, except for intermittent gaps in the data, the distribution of observations is uniform in time. There are only small variations in the number of observations per day. However, as a consequence of the Sun-synchronous orbit, the distribution of observations in local time is very uneven and, as shown in Figure 8, depends on the latitude. For observations made from Metop, the diurnal cycle is never fully sampled except near the poles.

Other RO missions, e.g. COSMIC and CHAMP, were designed to have orbital precession rates that let them drift slowly in local time such that over periods of months the full diurnal cycle is sampled. For example, the CHAMP orbit takes about 260 days to rotate through a full solar day. This type of drift is schematically shown by the coloured lines in Figure 7.

These sampling distributions indicate two different strategies to overcome problems related to detection of long-term trends in climate data derived from observations that do not sample the full diurnal cycle: 1) Always observe the same local time and acknowledge that any trends are only valid for that local time, or combination of local times. 2) Sample the full solar day over a certain time period, which for CHAMP or GRACE would be a few months).

METOP Occultations: Local Time vs. Day Number Jul-Sep 2009
 - 3 months -

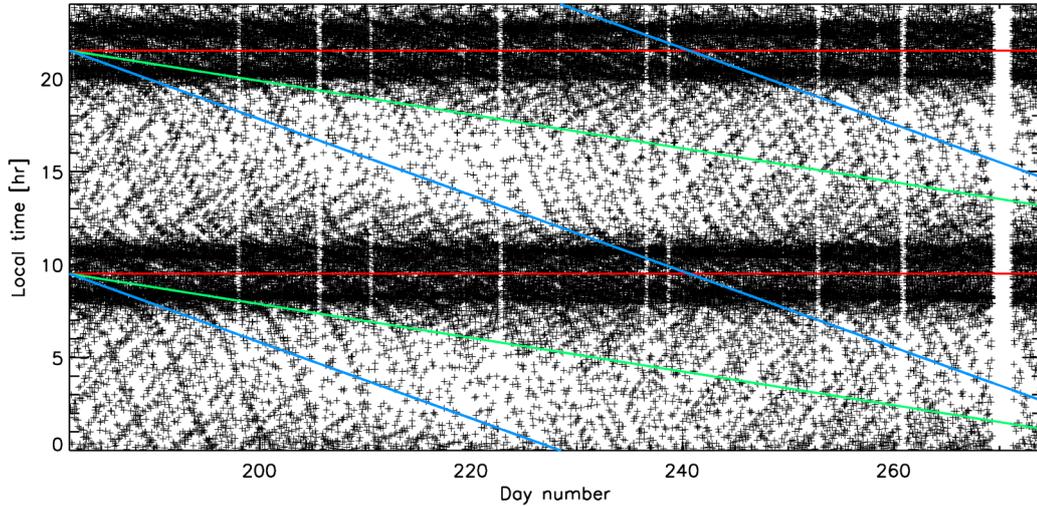
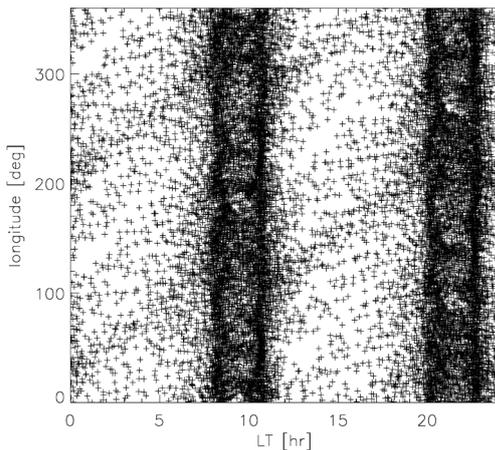


Figure 7. The local times of GRAS/Metop occultation events during the 3-month period July-Sep 2009. The coloured lines show the equatorial-crossing times for simulated LEO satellites with an approximate Metop orbit (red lines; inclination=98.7 deg), an approximate CHAMP orbit (green lines; inclination=87.2 deg), and an approximate COSMIC orbit (blue lines, inclination=72.0 deg). The different inclinations of the orbits, and to a lesser extent the other orbital parameters, give rise to different orbital precession rates, seen as a drift in local time.

METOP Occultations: Lon & LT Aug 2009
 - one month -



METOP Occultations: lat & LT Aug 2009
 - one month -

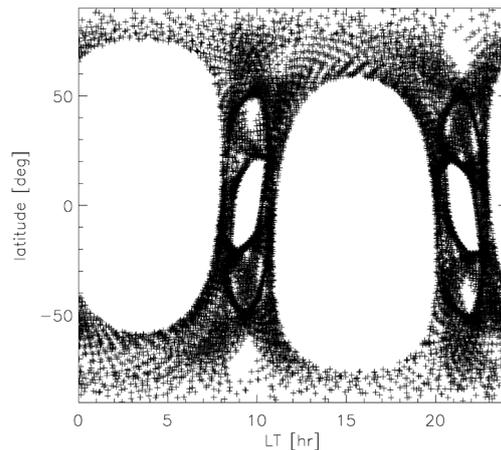


Figure 8. Scatter of observations during August 2009 in longitude and local time (left panel) and in latitude and local time (right panel). Since the Metop orbit is Sun-synchronous, these patterns are approximately stationary. For RO missions with orbits precessing relative to the Sun, the corresponding patterns would gradually move in local time.

4.6 Long-term stability of data

RO measurement data are very stable in time compared to most satellite- and ground-based atmospheric sounding techniques. The RO technique is sometimes described as “calibration-free”, meaning that in principle there is no need to calibrate an RO instrument to obtain long, homogeneous time series of data, or even to do inter-calibration between different instruments and missions [RD.16]. The reason for this is that the RO technique is based on measuring time differences – which are fundamentally SI traceable – rather than radiances, which makes it different from techniques based on passive infrared, visual, or microwave atmospheric sounding.

In the rather complex processing chain leading from fundamental observables to bending angle and refractivity there are, however, other potential causes of biases that may, or may not, be time varying. The algorithms and software – from the firmware operating the instrument, through all intermediate processing steps, to the final inversion to geophysical variables – are likely to evolve in time. This requires documentation and strict procedures for software updates. A more important source of bias is time varying noise of instrumental or ionospheric origin, which may have an impact on the retrieved geophysical variables by shifting the amount of background (*a priori*) information. There are many possible causes of such noise and many ways in which the noise characteristics may change. This requires monitoring of relevant quantities. In Annex I it is described how one can monitor noise and errors in bending angle, in refractivity, and in the model background used as *a priori* in the 1D-Var retrieval of temperature and humidity.

5. Level 3 algorithms

5.1 Algorithms overview

The Level 3 gridded data are generated from the Level 1B and Level 2 profile data through rather straight-forward *binning and averaging*. A set of equal-angle latitudinal bins are defined and all valid observations that fall within a latitude bin and calendar month undergo a weighted averaging to form a zonal mean for that latitude and month. The weighting is done by dividing each latitude bin into two (or more) sub-bins, computing an average for each of these, and then computing the mean of the two (or more) averages weighted by the areas of the sub-bins. The purpose of the weighting is to more closely approximate an area-weighted average.

The sampling errors are estimated by sub-sampling an atmospheric model (currently, a ECMWF reanalysis) at the observed times and locations. Based on these estimates, we can perform a sampling-error correction, or adjustment, simply by subtracting the estimated sampling errors from the observed means. The errors remaining after the sampling-error correction are referred to as residual sampling errors.

The uncertainty of the monthly mean is estimated as a combination of the per-profile measurement uncertainties and the uncertainties due to the residual errors remaining after the sampling-error correction. In principle, there is also a structural uncertainty due to algorithmic choices and underlying processing assumptions, but these are not explicitly quantified by the ROM SAF Level 3 algorithms. However, the ROM SAF has participated in activities with the explicit purpose to quantify structural uncertainties by comparing independent processing of the same input data. Results from these studies have been published in the scientific literature [RD.28,29].

In summary, the RO Level 3 gridded data are generated by the following steps:

- 1) quality control and flagging of profiles that are identified as non-nominal ('bad')
- 2) vertical interpolation of profiles onto a regular Level 3 height grid
- 3) weighted averaging into monthly latitude bins
- 4) estimation of sampling errors in the monthly means
- 5) estimation of uncertainties (measurement and sampling) in the monthly means
- 6) estimation of *a priori* information in the monthly means
- 7) formatting of the Level 3 gridded data and meta-data into netCDF files

The generation of zonally gridded monthly mean data may be followed by further averaging into seasonal and annual means, and into regional, hemispheric, and global means.

5.2 From profiles to gridded monthly mean data

5.2.1 Quality control of profiles

The purpose of the quality control is to identify profiles that are likely to provide an invalid representation of the true atmosphere. Before processing the atmospheric profiles into gridded monthly-mean data, all profiles are checked against a set of criteria indicating non-nominal conditions. Some of these criteria are seldom or never met – they are only a basic sanity check to ensure that corrupt data do not affect the climate data. When an occultation does not pass a test, the whole profile is discarded. No attempt is made to identify “good” data points within a profile containing “bad” data points.

The first step (*QC-0*) in the quality screening procedure is a basic check to ensure that the bending angle (refractivity) profile reaches above 60 km and below 20 km impact altitude (MSL altitude). Bending angles must fall within the range -1 to 100 mrad, and refractivities must fall within the range 0 to 500 N-units. The independent variables (impact altitudes and MSL altitudes) are required to vary monotonously.

In the next step (*QC-2*), the noise properties of the L2 signal and the degree of fit of the raw LC bending angle to the background bending angle is checked. The *L2 quality score* quantifies the degradation of the L2 signal through the RMS difference of the L1 and L2 impact parameter series obtained from a radio-holographic analysis [RD.8]. The two *SO scaling factors* quantify the degree of fit to a background bending angle profile [RD.2]. This QC step also includes a requirement that the background bending-angle data should only play a minor role below 40 km altitude, which is indicated by the *LC weighting factor* [RD.2]. This set of tests rejects around 5-10% of occultations, depending on which satellite mission it is.

The next QC step (*QC-3*) removes data identified as outliers. This is done by comparing the observed bending angles, refractivities, and dry temperatures to ECMWF reanalysis data. This set of tests rejects another 2-8% of occultations, depending on RO satellite mission.

If an occultation did not pass one or several of the above tests, the bending angle, refractivity, and dry variables are marked as non-nominal. Otherwise, they are regarded as nominal, and the refractivity profiles are passed on to the 1D-Var processing. This is followed by another QC step (*QC-4*) which checks the quality of the generated 1D-Var solution. Only a few percent of data are additionally rejected by this set of tests.

If the occultation passes all tests up to, and including, *QC-3*, but fails in *QC-4*, the bending angle, refractivity, and dry-variable profiles are used, while the wet profiles obtained from the 1D-Var solution are discarded.

<i>QC-0: basic sanity check</i>
<i>Identification of occultations with too small vertical extension, too few useful data points, the presence of invalid data points, or height variables that form a non-monotonous series.</i>
<ul style="list-style-type: none"> – $\alpha(H_a)$ must reach below 20 km and above 60 km – $\alpha(H_a)$ values must fall within valid range: [-1,100] mrad – H_a values must form a monotonous series – $N(H)$ must reach below 20 km and above 60 km – $N(H)$ values must fall within valid range: [0,500] N-units – H must form a monotonous series
<i>QC-1: (not used)</i>
<i>QC-2: bending angle quality</i>
<i>Checking of a) the quality of the bending angles, as quantified by the noise on the L2 impact parameter series, b) the fit of the raw LC bending angle to a background bending angle profile, and c) that the background bending-angle data only play a minor role below 40 km altitude.</i>
<ul style="list-style-type: none"> – L2 quality score must be less than 30.0 – SO scaling factor 1 must fall in the interval [0.92,1.08] – SO scaling factor 2 must fall in the interval [0.60,1.40] – LC weighting factor must be larger than 0.90 below 40 km altitude
<i>QC-3: identification of outliers</i>
<i>Identification of outliers by comparing with ECMWF reanalysis data mapped to refractivity, bending angle, and dry temperature.</i>
<ul style="list-style-type: none"> – α must deviate from reanalysis by less than 90% between 10–40 km – N must deviate from reanalysis by less than 10% between 5–35 km – N must deviate from reanalysis by less than 20% below 5 km – T_{DRY} must deviate from reanalysis by less than 20 K between 30–40 km
<i>QC-4: quality of 1D-Var solution</i>
<i>Identification of occultations that have problems converging at an acceptable 1D-Var solution.</i>
<ul style="list-style-type: none"> – the 1D-var algorithm must converge within 25 iterations – the penalty function $2J/N_{obs}$ must be smaller than 5.0 at convergence

Table 5.1. Summary of the ROM SAF quality control of the Level 1 and 2 data used as input to the Level 3 processing. When an occultation does not pass a test, the whole profile is discarded. No attempt is made to identify “good” data points within a profile containing “bad” data points. QC-0 to QC-3 affect all variables, while QC-4 only affects the 1D-Var variables. QC-1 is currently not used operationally (only used experimentally for screening based on noise in excess phase time series).

5.2.2 Interpolation of profiles onto the Level 3 height grids

There are three different height grids used in the ROM SAF Level 3 gridded data processing (see Sections 3.2 and 3.3 for definitions of the heights variables): the impact altitude (H_a) grid, the mean-sea level altitude (H) grid, and the pressure height (H_p) grid. During the Level 3 processing all three grids range from 0 to 80 kilometers with a 200 meter vertical spacing, but are finally cut at an upper altitude limit according to the requirements in the PRD before being stored as climate data products.

While the gridded data are given on a regular 200-meter grid, the profile input data are given on a slightly non-uniform grid, which is different from profile to profile. Before averaging, each profile is therefore interpolated onto a regular 200-meter height grid.

- Bending angle, α , is log-linearly interpolated onto the impact altitude grid assuming that $\log(\alpha)$ varies linearly with H_a between observed data points.
- Refractivity, N , is log-linearly interpolated onto the MSL altitude grid assuming that $\log(N)$ varies linearly with H between observed data points.
- Dry temperature, T_{dry} , is linearly interpolated onto the MSL altitude grid.
- Dry pressure, p_{dry} , is log-linearly interpolated onto the MSL altitude grid.
- Dry geopotential height, Z_{dry} , is linearly interpolated onto the pressure height grid.
- 1D-Var temperature, T , is linearly interpolated onto the MSL altitude grid.
- 1D-Var specific humidity, q , is log-linearly interpolated onto the MSL altitude grid.

5.2.3 Averaging in grid boxes

The Level 3 zonal grid consists of 36 equal-angle latitude bands spanning the entire globe from South Pole to North Pole. All latitude bins have a width of 5 degrees. Each RO profile is assigned to a single latitude bin and calendar month based on the reference location and time of the occultation.

The gridded monthly-mean data are obtained by a weighted arithmetic averaging of all interpolated atmospheric profiles that fall within a latitude bin and calendar month. The purpose of the weighting is to reduce the effects of variations in sampling density across a spatial bin, in order to obtain a correctly area-weighted mean. Depending on the character of the spatial sampling, different weighting strategies may be appropriate.

The distribution of observations in longitude is nearly uniform (see Section 4.5) and need not be explicitly addressed by weighting of the observations. This is not the case for the latitude distribution. Under the assumption that the observations have a uniform occurrence probability per degree of latitude, the profiles could be weighted by a factor $\cos(\varphi)$, where φ is the latitude. Under that assumption, a cosine weighting corrects for the fact that within a grid box there are on average slightly more observations *per unit area* at higher latitudes than at lower latitudes due to the meridian convergence. However, if the sampling distribution is not uniform in this sense, a simple cosine weighting may not be appropriate, even though it is the preferred method for averaging data taken from a regular grid.

The plots in Section 4.5 indicate that the distribution of observations in latitude is far from uniform in the sense assumed above, particularly at high latitudes. We therefore choose not to use a simple cosine weighting. Instead, the effects of non-uniform sampling in latitude are reduced by subdividing each 5-degree latitude bin into two 2.5-degree sub-bins, designated by S (“south”) and N (“north”). All profiles falling within a sub-bin are averaged, and the “south” and “north” mean profiles are then in turn averaged, weighted by the surface area of the respective sub-bin.

This is a general method in the sense that it assumes very little about the sampling distribution. Simulation experiments on different type of latitudinal distributions show that, compared to cosine weighting, subgridding gives smaller sampling biases for typical distributions of occultation events [RD.12].

Hence, for each latitude bin and for each of the variables α , N , T_{dry} , p_{dry} , Z_{dry} , T , and q , two mean profiles are computed, one for the southern (S) and one for the northern (N) sub-bin.

$$\bar{X}_S(h) = \frac{1}{n_S} \sum_{i_S} X_{i_S}(h) \quad (21)$$

$$\bar{X}_N(h) = \frac{1}{n_N} \sum_{i_N} X_{i_N}(h) \quad (22)$$

where X designates an atmospheric variable, n_S and n_N are the number of profiles in the two sub-bins, and indices i_N and i_S loop over the profiles in the two sub-bins, respectively. The two

mean profiles are combined into a total average for the bin, by weighting with the surface area of the respective sub-bin.

$$\bar{X}(h) = \frac{1}{A_S + A_N} [A_S \cdot \bar{X}_S(h) + A_N \cdot \bar{X}_N(h)]$$

The independent variable h indicates that the variables are functions of height. The number of observations, n_S and n_N , are also functions of height, even though it is not explicitly shown in the expressions above.

Equations 21–23 can be generalized to more than two sub-bins by giving each data point, i , a weight, w_i , according to which sub-bin, s , it belongs to

$$w_i = \frac{A_s}{A} \frac{n}{n_s} \quad (24)$$

where A and n are the total area and data number for the bin, and A_s and n_s are the area and data number for sub-bin s . The dependency on height, h , is not shown explicitly. Within each latitude bin, a weighted arithmetic mean can be computed according to

$$\bar{X}(h) = \frac{\sum w_i X_i(h)}{\sum w_i}$$

where

$$\sum w_i = n \quad (26)$$

The weighting described by Equations 21–23, or alternatively by equations 24–25, addresses the non-uniform *spatial* sampling that may occur in the grid boxes. The non-uniform *temporal* sampling described in Section 4.4 is not explicitly addressed by the averaging procedure. However, in principle, the weighting could also be applied to, e.g., longitude, local time of day, or day of the month.

The corresponding weighted standard deviation is given by

$$s(h) = \sqrt{\frac{\sum w_i (X_i(h) - \bar{X}(h))^2}{((n-1)/n) \sum w_i}} \quad (27)$$

using the same weights as in equations 24–25.

5.2.4 Estimating uncertainty in the means

The *observed grid-box mean* described in Section 5.2.3 is an estimate of the *true grid-box mean*. The difference between the observed mean and the true mean is referred to as the *error* of the climate variable. This error is assumed to be caused by two effects. First, each measurement has a random *measurement error* associated with it. This error can only be described in terms of a statistical uncertainty. Secondly, the finite number of measurements is not able to fully account for all variability within the latitude bin and time interval, resulting in a *sampling error*. Unlike the measurement errors, it is possible to estimate the actually realized sampling errors in the monthly means. This allows us to make a correction of the observed means, leaving a *residual sampling error*. We assume the residual sampling error to be random, and we describe it in terms of a statistical uncertainty.

The *measurement uncertainty of the mean* is estimated by assuming that the n data points, i , within a latitude bin have associated random measurement uncertainties $\sigma_{i,\text{meas}}$. The assumed measurement uncertainties for the profiles are described in Section 4.4. Assuming that all observations are independent, the two sub-bin means have estimated uncertainties $\sigma_{\text{S},\text{meas}}$ and $\sigma_{\text{N},\text{meas}}$ given by

$$\begin{aligned}\sigma_{\text{S},\text{meas}}^2 &= \frac{1}{n_{\text{S}}^2} \sum_{i_{\text{S}}} \sigma_{i_{\text{S}},\text{meas}}^2 \\ \sigma_{\text{N},\text{meas}}^2 &= \frac{1}{n_{\text{N}}^2} \sum_{i_{\text{N}}} \sigma_{i_{\text{N}},\text{meas}}^2\end{aligned}\quad (29)$$

where n_{S} and n_{N} are the number of observations in the two sub-bins, and indices i_{N} and i_{S} loop over the data in the two sub-bins, respectively. The two uncertainties are combined into an estimated uncertainty, σ_{meas} , in the total mean:

$$\sigma_{\text{meas}}^2 = \frac{1}{(A_{\text{S}} + A_{\text{N}})^2} \left[A_{\text{S}}^2 \cdot \sigma_{\text{S},\text{meas}}^2 + A_{\text{N}}^2 \cdot \sigma_{\text{N},\text{meas}}^2 \right]$$

where A_{S} and A_{N} are the areas of the two sub-bins and A is their sum.

Similar to the averaging (Section 5.2.3), equations 28–30 can be generalized to more than two sub-bins by giving each data point, i , a weight, w_i , according to equation 24. The estimated measurement uncertainty, σ_{meas} , in the mean is then given by

$$\sigma_{\text{meas}}^2 = \frac{\sum w_i^2 \sigma_{i,\text{meas}}^2}{\left(\sum w_i\right)^2}$$

where index i loops over all data in the bin.

The *sampling error of the mean* is estimated by sampling an atmospheric model at the same times and locations as the observations. In the ROM SAF Level 3 processing, sampling errors are estimated from ECMWF reanalysis fields at a coarse 2.5°x2.5° horizontal resolution. This grid has been interpolated from a spectral model truncated to T63, which has a horizontal resolution comparable to the observations.

The full-grid model means are computed from the 4-dimensional model field

$$\bar{X}_{\text{model}}^{\text{grid}} = \frac{1}{n_t n_\varphi n_\lambda} \frac{1}{\sum_k \cos(\varphi_k)} \sum_{t=1}^{n_t} \sum_{k=1}^{n_\varphi} \sum_{l=1}^{n_\lambda} X_{tkl} \cdot \cos(\varphi_k) \quad (32)$$

using the cosine weighting discussed in Section 5.2.3. The summation in Eq. 32 loops over all n_t , n_φ , and n_λ time-latitude-longitude model grid points located within the climate data grid box. The dependency on height, h , is not shown explicitly in Eq. 32. For data on a 2.5-degree latitude grid, Eq. 32 gives the same results as Eqs. 21 to 23.

Under the assumption that the model variability is statistically representative for the true atmosphere, the difference between the sub-sampled means, $\bar{X}_{\text{model}}^{\text{col}}$, computed from model data co-located with the observations, and the full-grid means

$$\Delta\bar{X} = \bar{X}_{\text{model}}^{\text{col}} - \bar{X}_{\text{model}}^{\text{grid}} \quad (33)$$

provides an estimate of the sampling error of the mean

$$\varepsilon_{\text{samp}} \approx \Delta\bar{X} \quad (34)$$

This estimate allows us to make a correction, or adjustment, of the observed means by subtracting the estimated sampling error from the observed mean, leaving a residual sampling error, $\varepsilon_{\text{resamp}}$. Studies [RD.21] show that this correction leaves around 20% to 30% of the sampling error, which can be described in terms of a quasi-random *residual sampling uncertainty*, σ_{resamp} .

The measurement uncertainty, σ_{meas} , and the residual sampling uncertainty, σ_{resamp} , are independent and can be combined into a total uncertainty for the mean

$$\sigma_{\text{clim}} = \sqrt{\sigma_{\text{meas}}^2 + \sigma_{\text{resamp}}^2} \quad (35)$$

5.2.5 Estimating *a priori* information in the means

The grid-box means described in Section 5.2.3 are not a result of observational data alone, but also depend on *a priori* data – background atmospheric data taken from a model. This was mentioned in Section 4.1. There are two sources of *a priori* information: (a) background bending angle profiles that are used to smooth the observed bending angles and extend them to infinity before inverting bending angles to refractivity, and (b) the background atmospheric states that are used to resolve the temperature-humidity ambiguity through a 1D-Var procedure.

From Equation 4 we define the parameter

$$W_{\text{SO}} \equiv 1 - \frac{\sigma_{\text{bg}}^2}{\sigma_{\text{bg}}^2 + \sigma_{\text{obs}}^2} = \frac{\sigma_{\text{obs}}^2}{\sigma_{\text{bg}}^2 + \sigma_{\text{obs}}^2} \quad (36)$$

where σ_{obs} and σ_{bg} are estimates of the errors in the observed and background bending angles. The parameter W_{SO} provides a measure of the observational information in the optimized bending-angle profile. As a consequence of the error characteristics, W_{SO} is a monotonic function of altitude and goes from 0 (no background data) at low altitudes to 1 (no observational data) at high altitudes.

We now define a corresponding quantity for a grid-box mean

$$\bar{W}_{\text{SO}} \equiv \frac{1}{n} \sum_i \frac{\sigma_{\text{obs},i}^2}{\sigma_{\text{bg},i}^2 + \sigma_{\text{obs},i}^2} \quad (37)$$

using the same notation as in Section 5.2.3: index i loops over all data in the grid box.

Similarly, the observation and background error covariances assumed in the 1D-Var retrieval (Section 4.1) provide a means to quantify the relative importance of the *a priori* and observational information in the temperature and humidity data. Following the notation in Section 4.1, we let \mathbf{O} and \mathbf{B} denote the observational and background error covariance matrixes, respectively. The solution error covariance matrix \mathbf{S} is then given by

$$\mathbf{S}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{O}^{-1} \mathbf{H} \quad (38)$$

in the linear limit, i.e. in the limit where the forward model $H(\mathbf{x})$ can be adequately represented by its Jacobian \mathbf{H} . The second term on the right hand side is the observational error covariances forward modelled into the background state space.

Following Rieder and Kirchengast [RD.17] we let the error standard deviations – i.e. the square root of the diagonal elements of the error covariance matrixes – quantify the relative importance of the *a priori* information in the retrieved temperature and humidity:

$$W_T = 100 \cdot \frac{\sigma_{T,\text{sol}}}{\sigma_{T,\text{bg}}}$$

and

$$W_q = 100 \cdot \frac{\sigma_{q,\text{sol}}}{\sigma_{q,\text{bg}}} \quad (40)$$

where index ‘sol’ denotes solution and index ‘bg’ denotes background. The factor 100 normalizes the ratio to percent.

5.3 Time series and anomaly data

The gridded monthly-mean climate data records are fundamentally 3-dimensional: time series of zonal monthly means on a 2D latitude-height grid

$$f_{ijt} = f(\varphi_i, h_j, t) \quad (41)$$

where f is a climate variable (refractivity, temperature, etc.), indices i and j denote the latitude and height bins, and t denotes the time (here, month number).

We define the *long-term climate mean* as the mean over the time dimension for the full length of the time series

$$f_{ij}^C = f^C(\varphi_i, h_j) = \frac{1}{N_{\text{mon}}} \sum_{t=1}^{N_{\text{mon}}} f_{ijt} \quad (42)$$

where N_{mon} is the number of months in the climate data record. The long-term climate mean is used as a reference for constructing *anomalies* in which the seasonal cycle is retained.

We define the *mean annual cycle* as

$$f_{ijs}^{AC} = f^{AC}(\varphi_i, h_j, s) = \frac{1}{N_{\text{yr}}} \sum_{k=1}^{N_{\text{yr}}} f_{ijt} \quad (43)$$

where

$$t = 12 * (k - 1) + s \quad (44)$$

and s denotes the season (1 to 12) and N_{yr} is the number of years in the climate data record. The mean annual cycle is used to construct *de-seasonalized anomalies*, i.e. anomalies with the dominating seasonal cycle removed.

Based on the long-term climate mean we define the *anomalies* as

$$\Delta f_{ijt} = f_{ijt} - f_{ij}^C \quad (45)$$

and the *fractional anomalies* as

$$\Delta f_{ijt} = (f_{ijt} - f_{ij}^C) / f_{ij}^C \quad (46)$$

where the latter are used for quantities that have a predominantly exponential altitude dependence.

Similarly, based on the mean annual cycle we define the *de-seasonalised anomalies* as

$$\Delta f_{ijt}^{\text{des}} = f_{ijt} - f_{ijs}^A \quad (47)$$

and the *de-seasonalised fractional anomalies* as

$$\Delta f_{ijt}^{\text{des}} = (f_{ijt} - f_{ijs}^A) / f_{ijs}^A \quad (48)$$

where s is the season (1,...,12) for month t .

The anomaly fields are still 3D as they depend on latitude, height, and time.

For plotting, the number of dimensions needs to be reduced. This is done by averaging over a *latitude band* and *height layer*, which covers several bins. Averaging over a latitude band (considering the area of the latitude bins) gives a 2D time-height plot, averaging over a height layer gives a 2D time-latitude plot, and averaging over both a latitude band and a height layer gives a 1D time series plot.

6. Level 3 gridded data products

6.1 Product types: offline data, CDRs, and ICDRs

As described in the definitions section (Section 1.4), the CDRs have been generated in a dedicated reprocessing activity using the same algorithms throughout the length of the data records, while ICDRs are generated on a regular basis with the same algorithms as the CDRs, but using currently available input data. The main rationale for the ICDRs is that they extend the CDRs until data from a new reprocessing become available. There is a strong focus on the consistency between the ICDRs and the CDRs. In addition, the ROM SAF data product portfolio also includes offline data that are generated on a regular basis for non-time-critical applications, based on algorithms that may have evolved somewhat from the last reprocessing. All the ROM SAF Level 3 gridded data – irrespective of whether they belong to an offline data set, a CDR, or an ICDR – contain the same set of geophysical variables, are similarly time averaged and gridded, and use the same file formats.

6.2 Geophysical variables and height variables

The ROM SAF monthly mean gridded data include the following geophysical variables:

- Bending angle [mrad]
- Refractivity [N-units]
- Dry temperature [K]
- Dry pressure [hPa]
- Dry geopotential height [m]
- Temperature [K]
- Specific humidity [g/kg]
- Tropopause height [m]

The corresponding height variables are *impact altitude* for bending angles, *dry-pressure height* for dry geopotential height, and *mean-sea level (MSL) altitude* for the other variables. The impact altitude and the MSL altitude are referenced to the Earth's geoid, while the pressure height is a logarithmic measure of the pressure. The height variables are discussed in more detail in [RD.6].

6.3 Data grids

The geophysical variables are provided as monthly means on 2D latitude-height grids with a resolution of 5 degrees (in latitude) by 200 meters (in height). There is no longitudinal dimension – though formally there is a longitude dimension with size one. The type of vertical grid (impact altitude, mean-sea level altitude, or pressure height) depends on the geophysical variable (see Sections 4.1 and 4.2).

6.4 Numerical data

For each geophysical variable, the following numerical data are included as a part of the ROM SAF Level 3 gridded data product:

- monthly means (sampling-error corrected)
- monthly standard deviations
- estimate of sampling errors in the monthly means
- estimate of measurement uncertainty in the monthly means
- data number per bin
- monthly means of co-located ECMWF reanalysis data
- monthly standard deviations of co-located ECMWF reanalysis data

The refractivity, temperature, and specific-humidity data products also include:

- measure of *a priori* information content in the monthly means

6.5 Example data

The algorithms described in Section 5 and in Annex I have been extensively tested on RO data from the CHAMP, GRACE, COSMIC, and Metop missions. We here present a selection of Level 3 gridded data that were generated using the algorithms described in the preceding sections. Most of the examples are taken from the ROM SAF reprocessed Level 3 data sets generated from Metop data. The reference data referred to as “BGR” is ERA-Interim short-term forecasts. A few examples in Annex I are taken from CHAMP data sets used for testing. For the calculation of the residual sampling uncertainty in the plots, we used 30% of the estimated sampling errors (see Section 5.2.4).

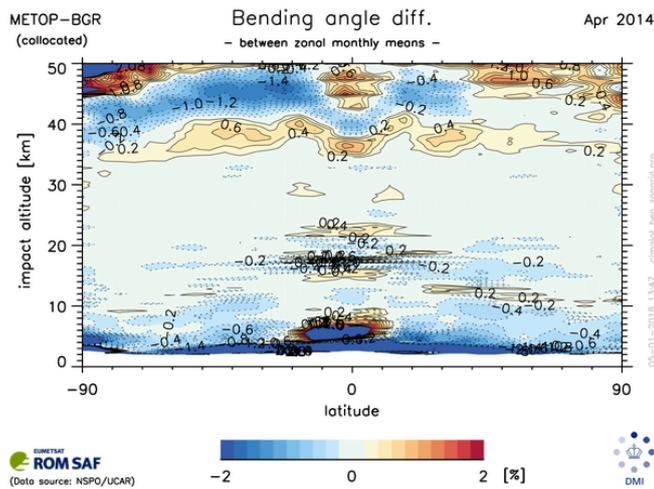
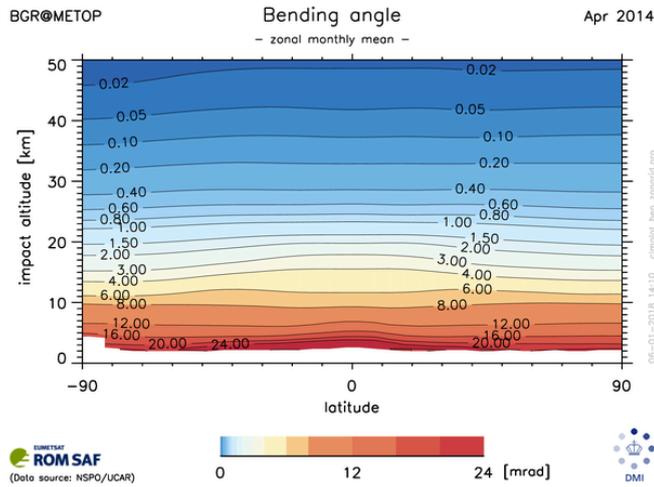
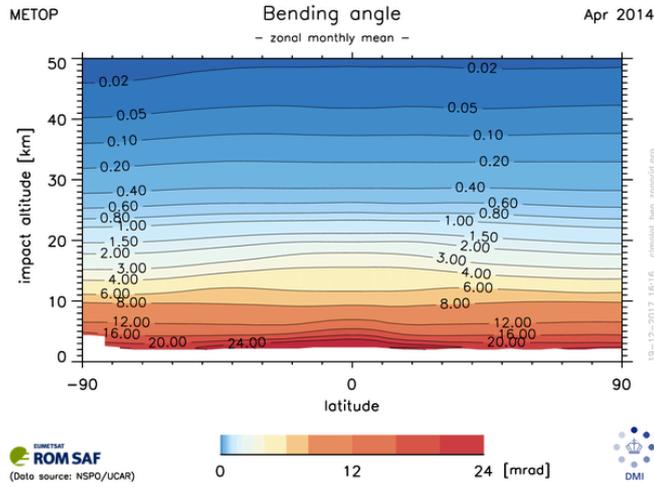


Figure 9a. Zonal monthly mean bending angle for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

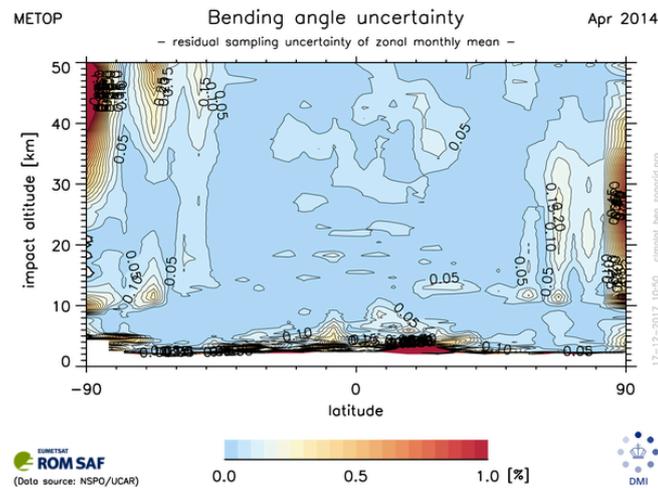
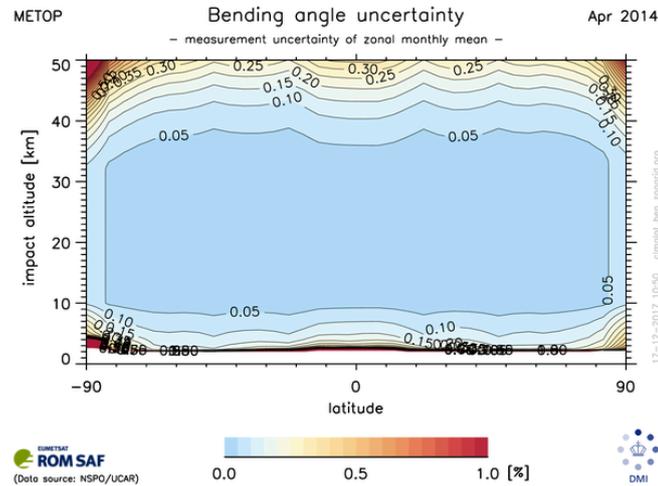
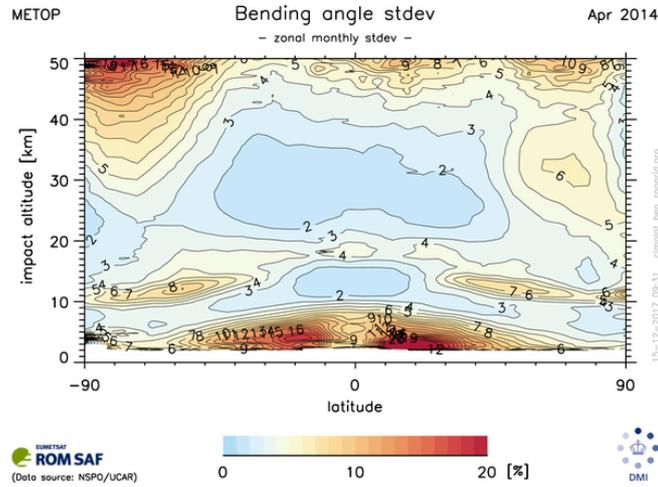


Figure 9b: The upper panel shows bending angle zonal monthly standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

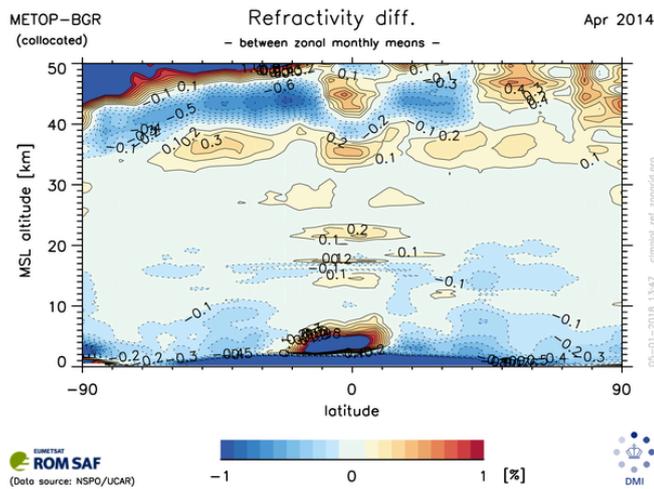
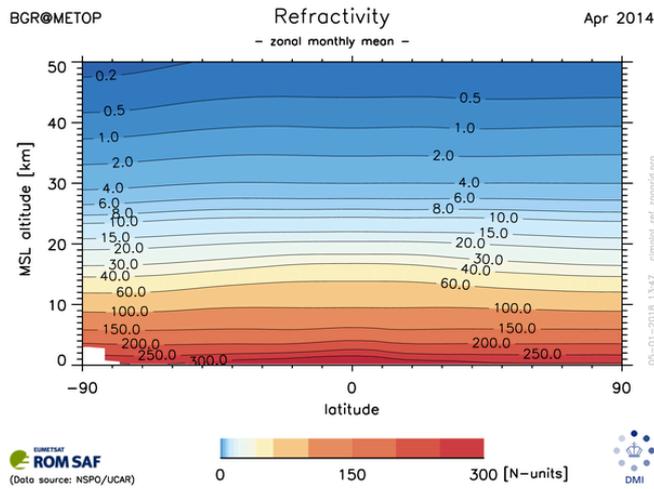
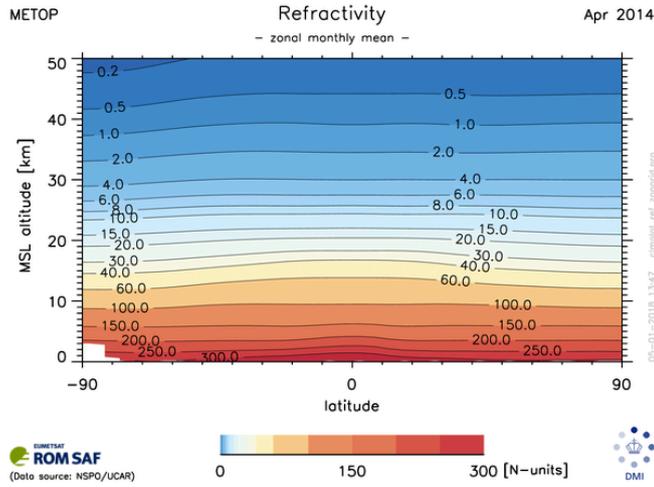


Figure 10a. Zonal monthly mean refractivity for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

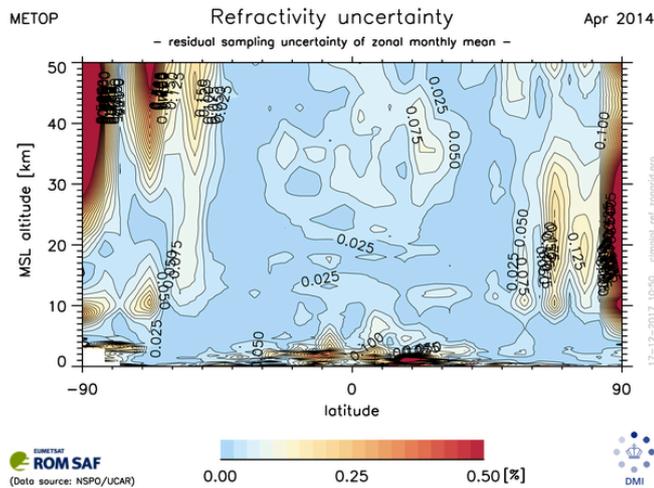
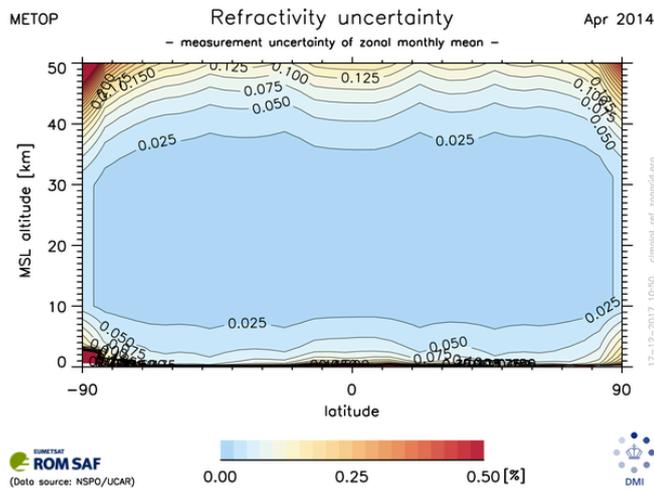
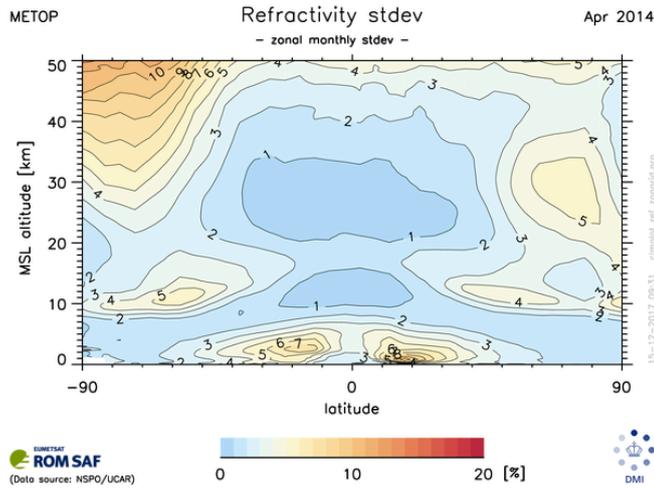


Figure 10b: The upper panel shows refractivity zonal monthly standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

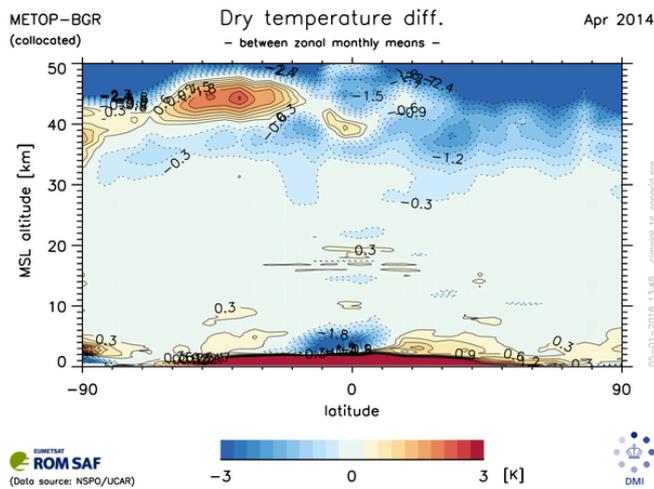
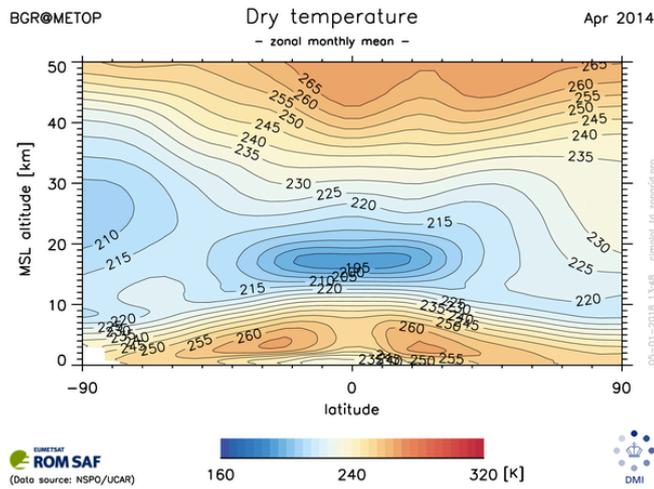
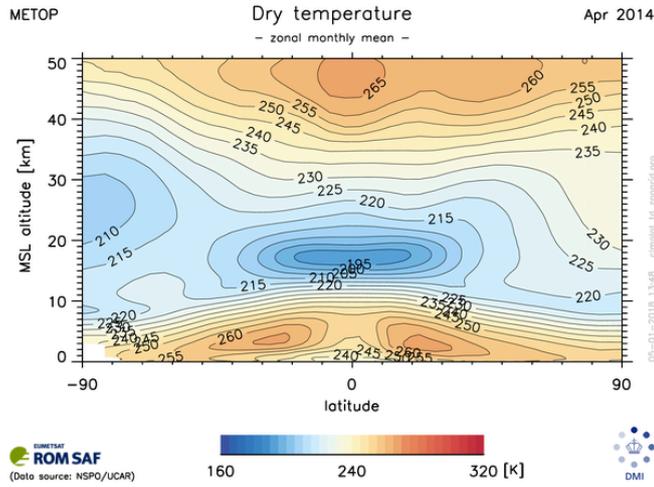


Figure 11a. Zonal monthly mean dry temperature for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

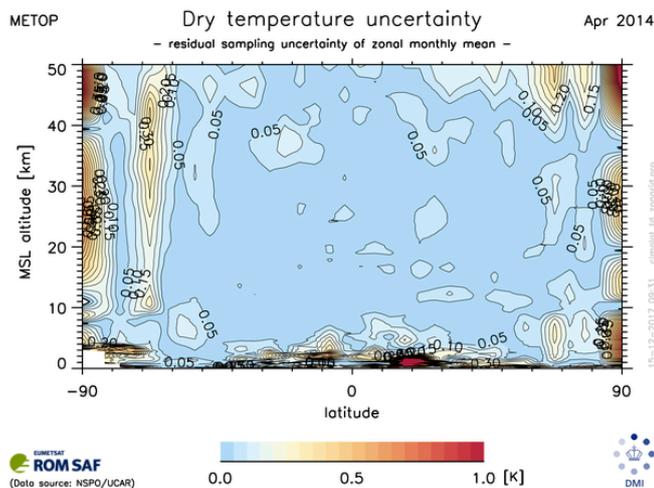
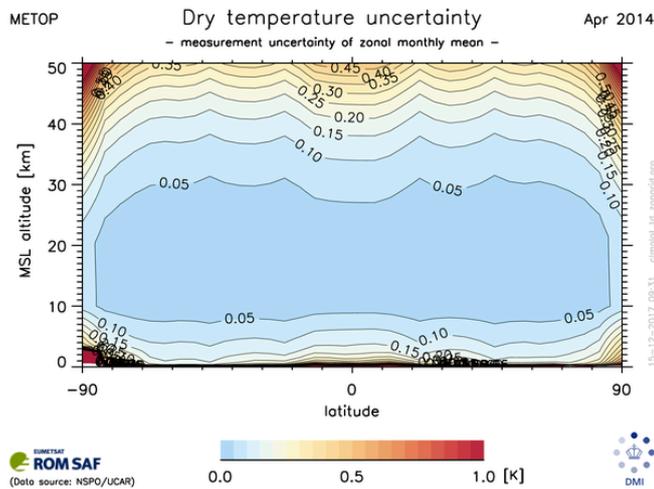
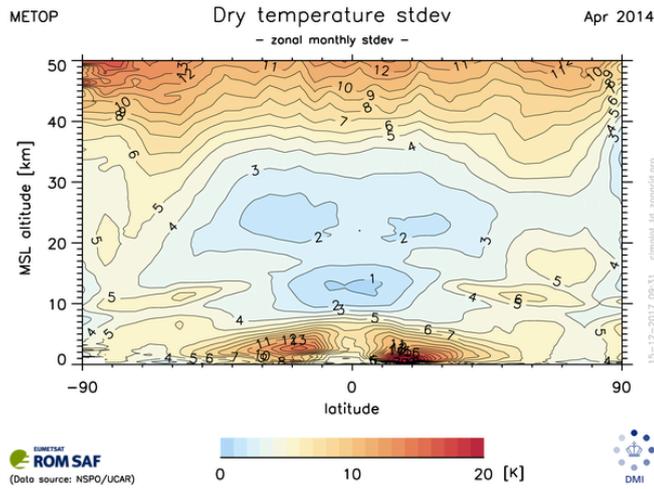


Figure 11b: The upper panel shows dry temperature zonal monthly standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

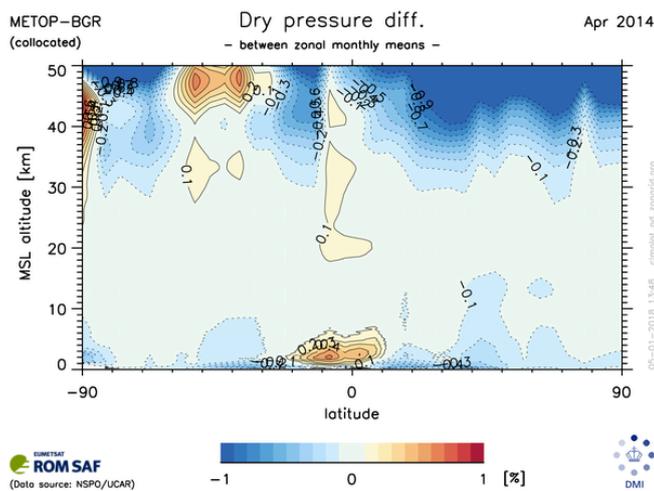
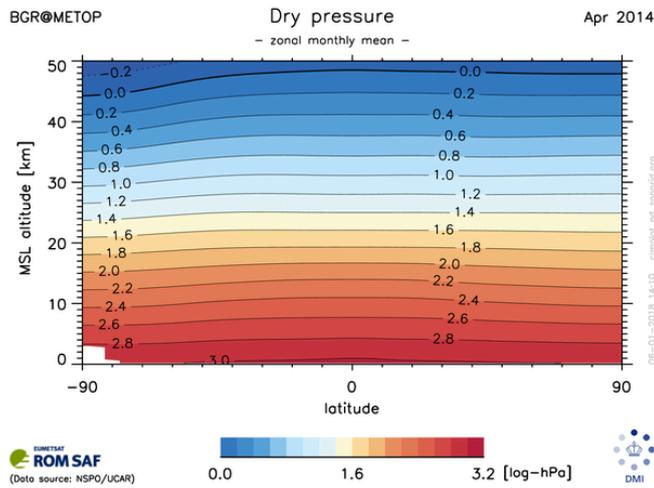
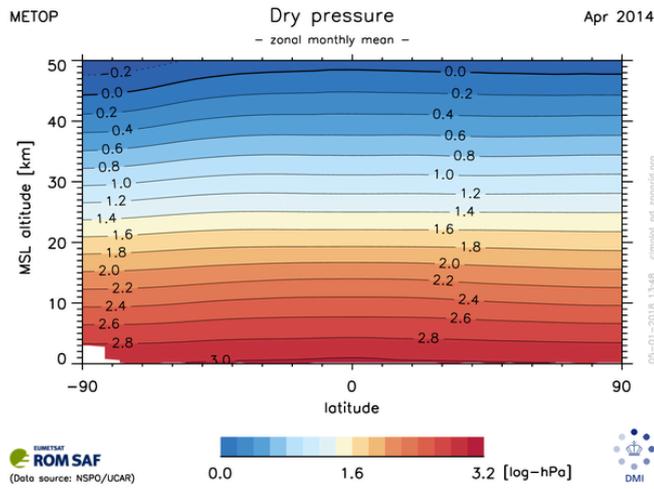


Figure 12a. Zonal monthly mean dry pressure for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

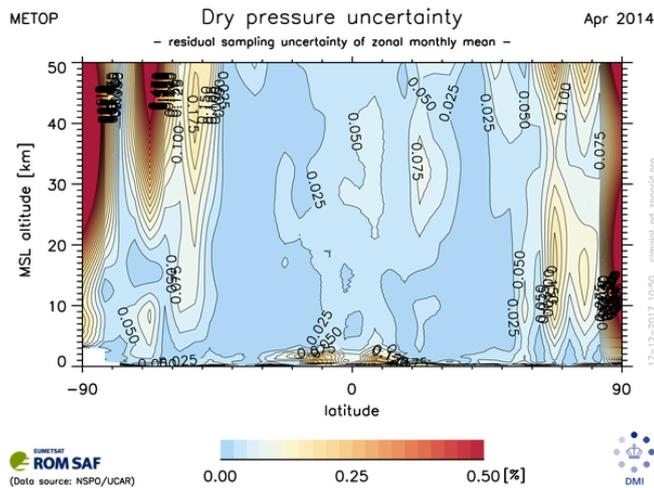
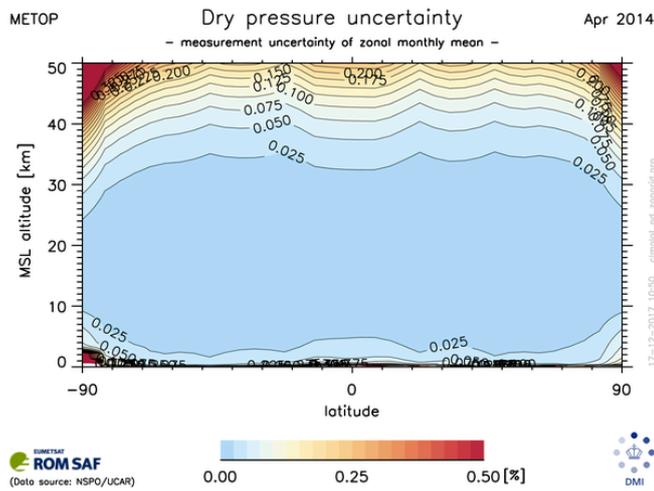
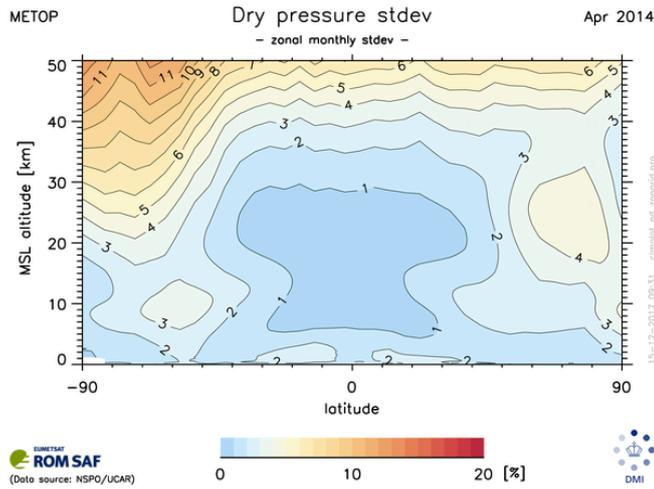


Figure 12b: The upper panel shows dry pressure zonal monthly standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

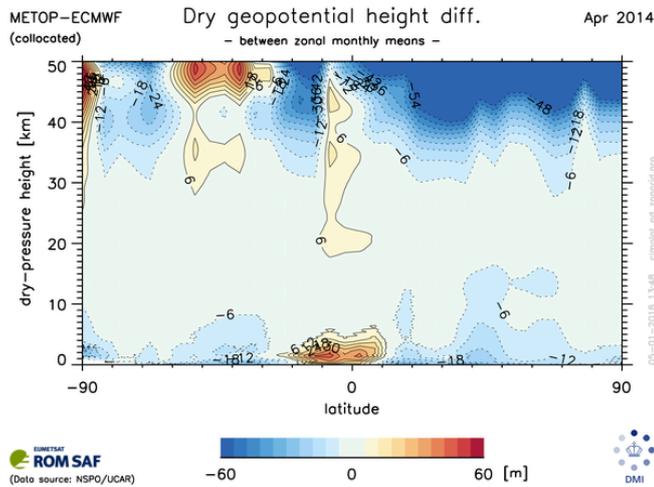
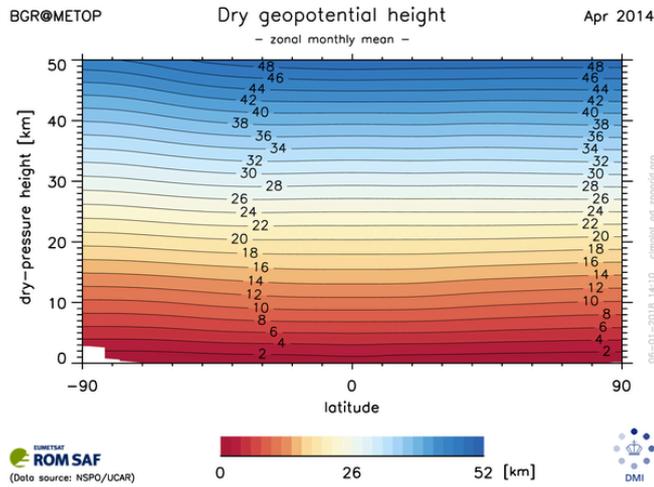
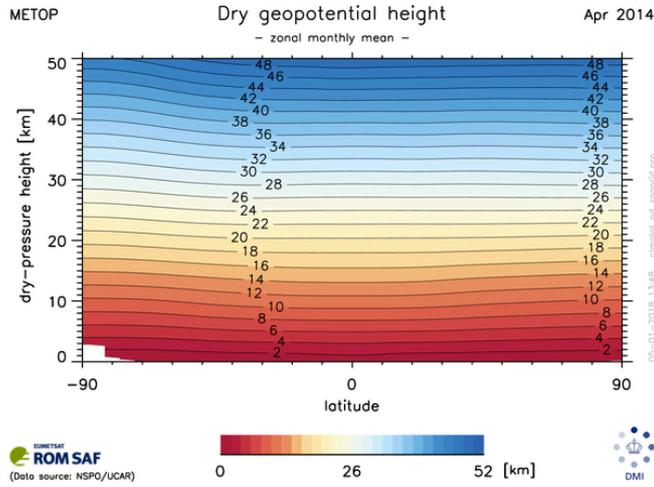


Figure 13a. Zonal monthly mean dry geopotential height for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

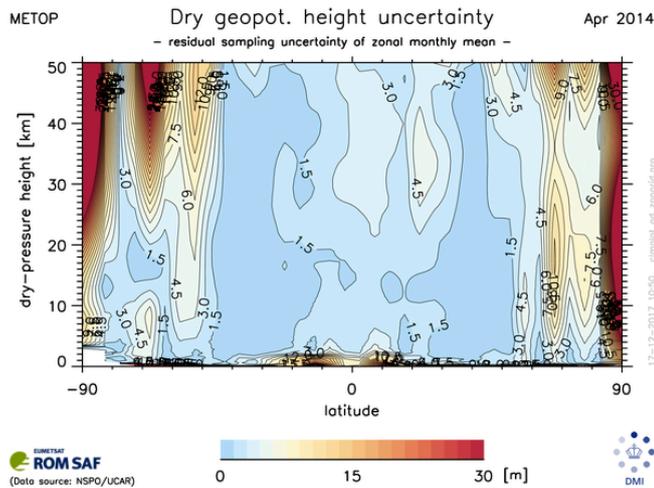
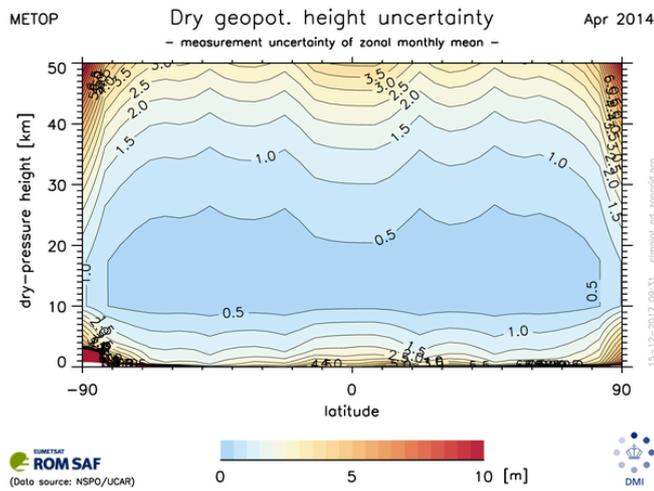
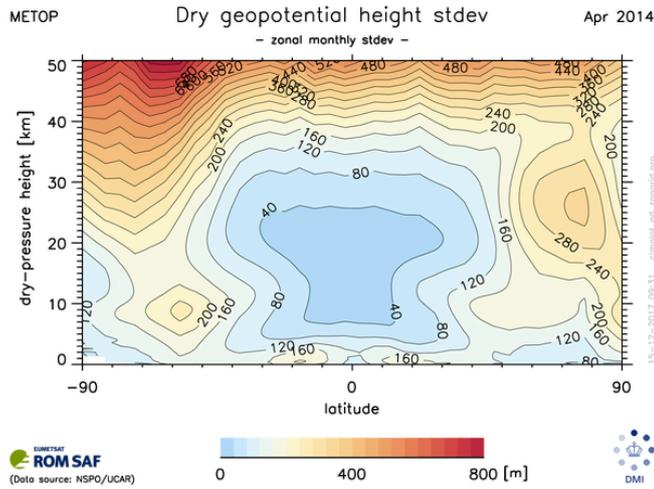


Figure 13b: The upper panel shows dry geopotential height zonal monthly standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

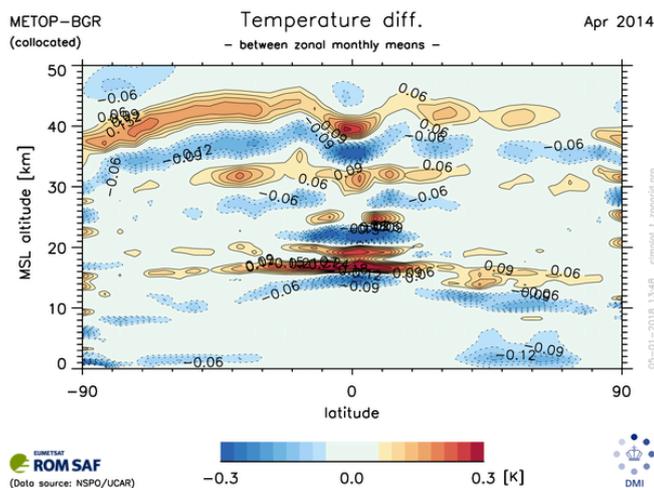
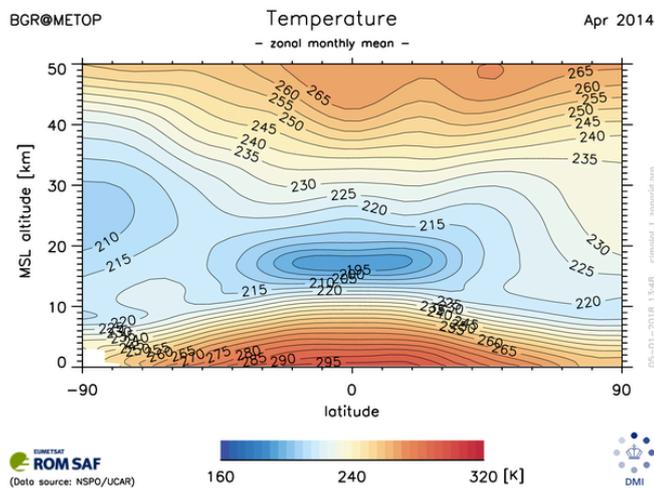
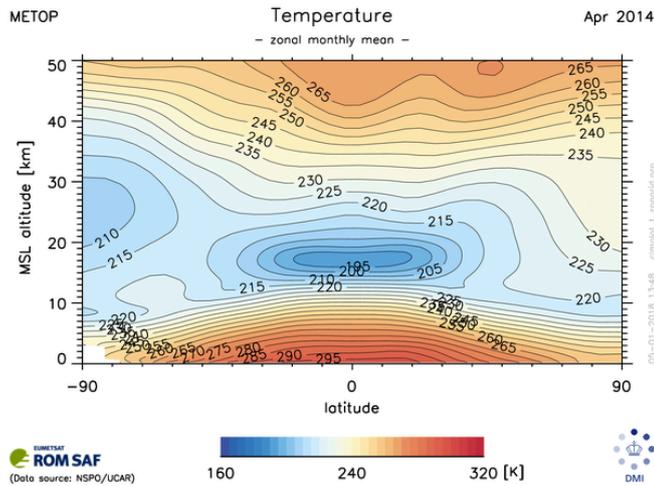


Figure 14a. Zonal monthly means of 1D-Var temperature for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

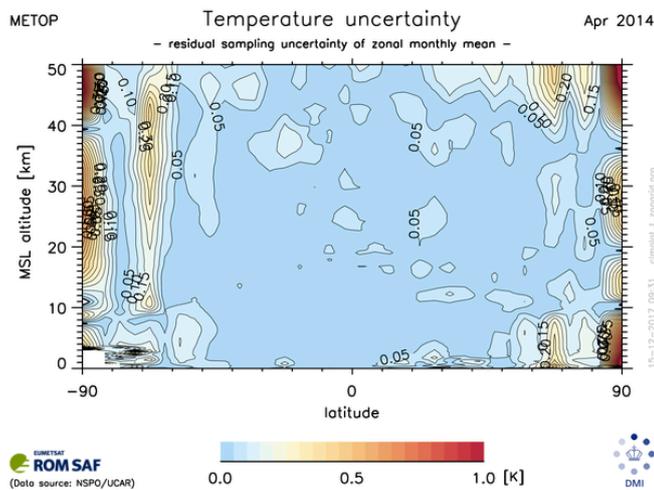
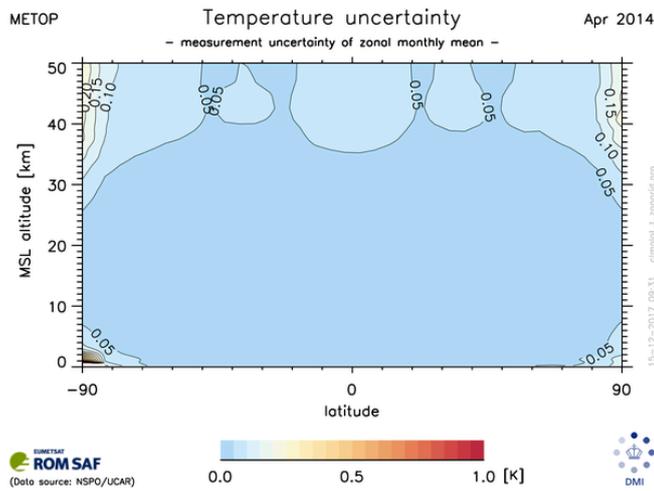
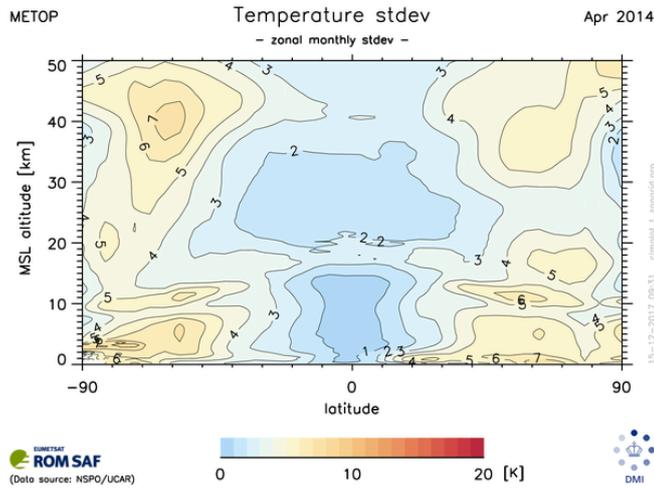


Figure 14b: The upper panel shows monthly temperature standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

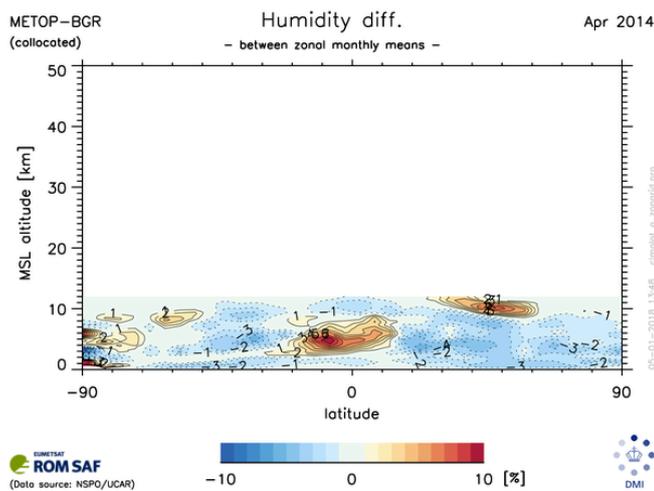
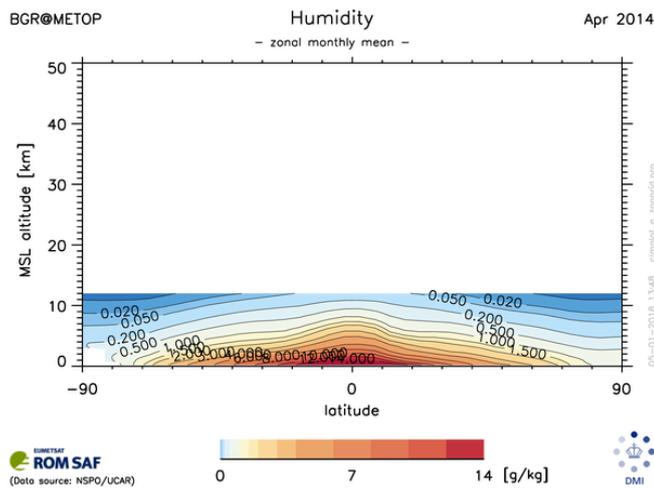
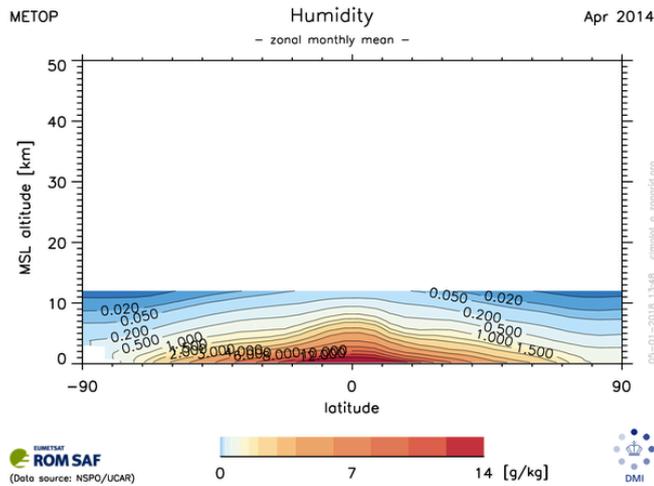


Figure 15a. Zonal monthly means of 1D-Var humidity for observed Metop data and for ERA-Interim short-term forecast data co-located with the observations. The bottom panel shows the differences between observation and ERA-Interim.

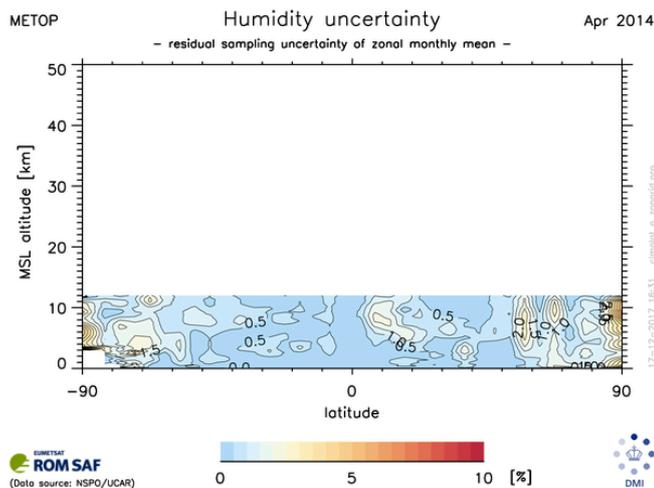
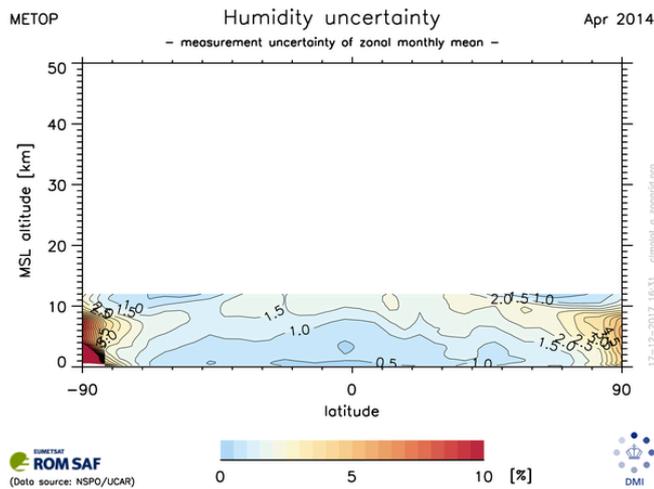
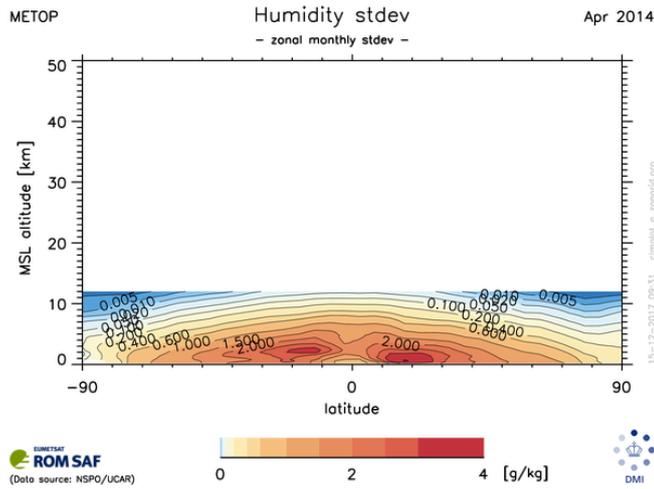


Figure 15b: The upper panel shows monthly humidity standard deviations for April 2014, based on data from the Metop mission. The middle and lower panels show estimates of the measurement uncertainty and the residual-sampling uncertainty of the monthly means.

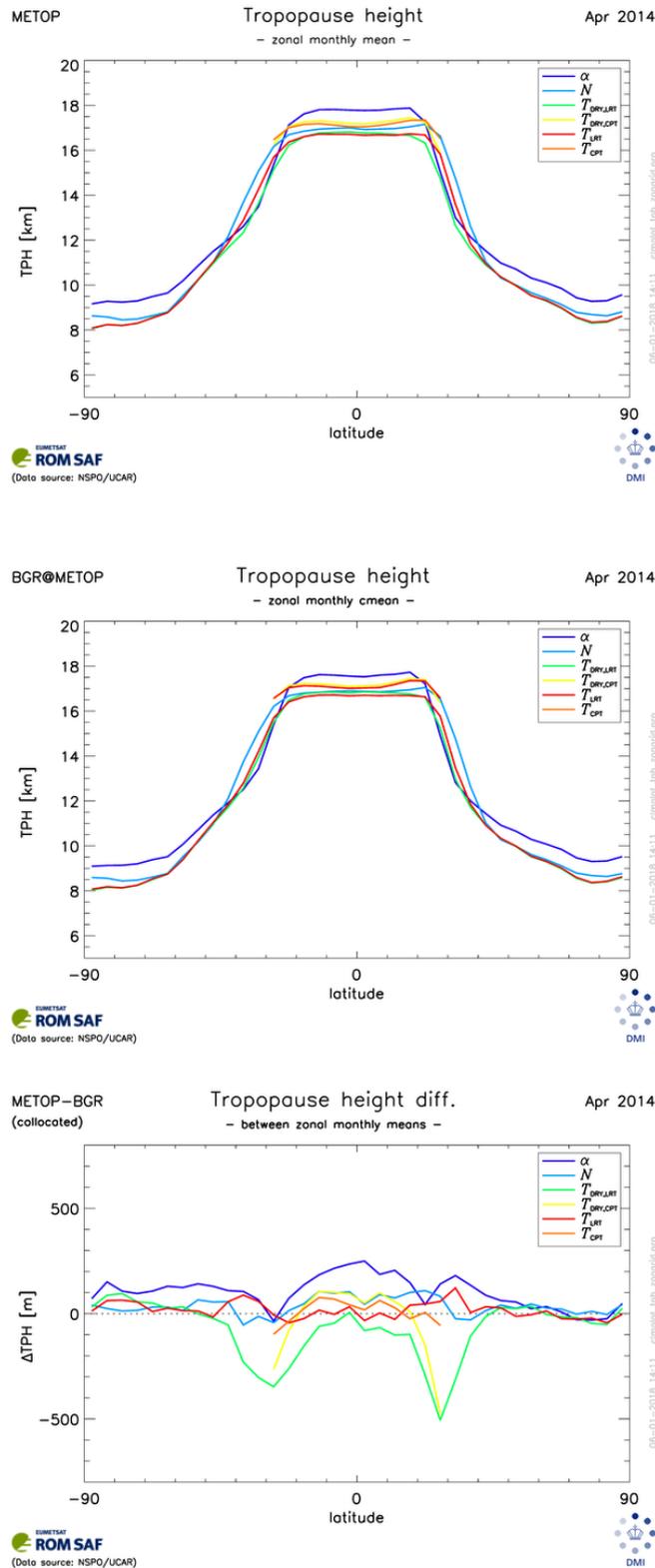


Figure 16. Zonal monthly means of the tropopause for observed Metop data and for ERA-Interim short-term forecast data collocated with the observations. The bottom panel shows the differences between observation and ERA-Interim. It is the tropopause based on the dry temperature lapse rate (green line) that is the formal ROM SAF product.

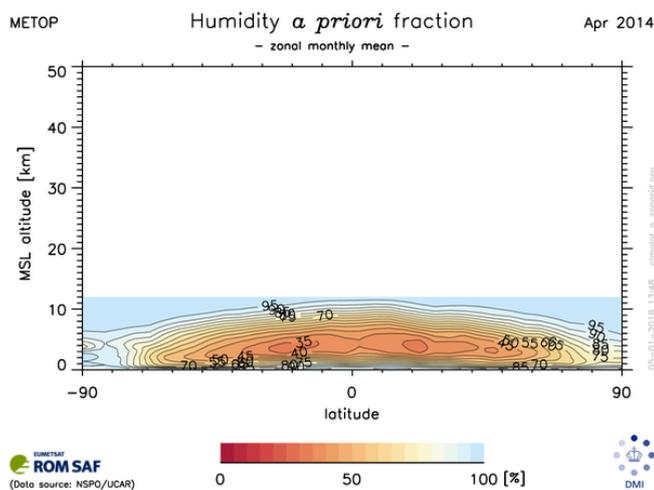
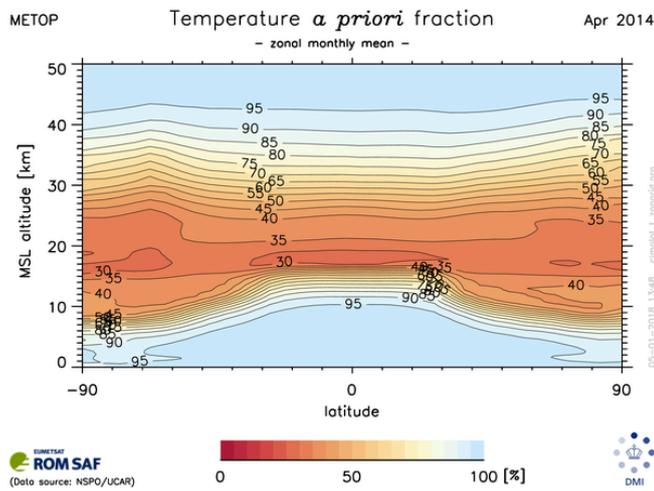
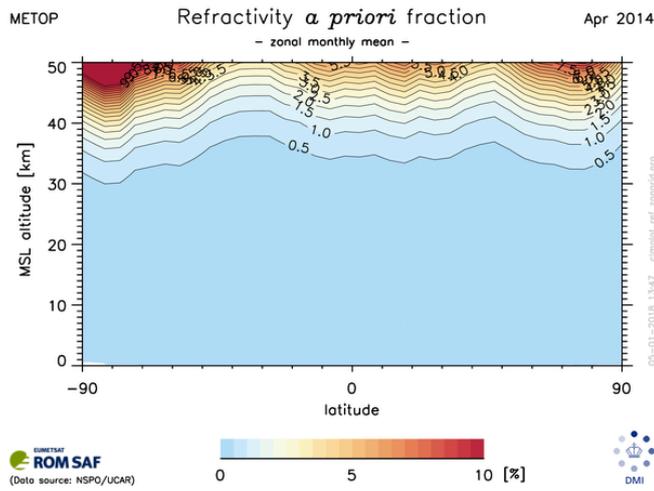


Figure 17. Relative importance of *a priori* in zonal monthly mean refractivity, temperature, and humidity for the Metop mission for April 2014.

Annex I: Monitoring long-term stability of the climate data

Climate data records require not only that random and systematic errors are small, but also that the errors do not change in time. The latter is referred to as stability – or long-term stability when stability is required over time periods of years and decades.

The primary means to assess whether data are stable is to monitor errors and/or noise in the data over long time periods. For this purpose we have defined a set of monitoring quantities. Any significant variations in time of these quantities should call for investigations concerning the causes of the variations – whether they are instrumental, or traceable to the processing system or changing conditions in the atmosphere or ionosphere.

I.1 Bending angle noise

The influence of the ionosphere on the bending angle is removed to first order by forming a linear combination of the L1 and L2 bending angles. The neutral-atmosphere bending angle (the “raw” bending angle) thus obtained is contaminated with noise that increases exponentially with altitude. This noise is of both instrumental and ionospheric origin and varies considerably from occultation to occultation.

We estimate the upper level bending-angle noise σ_α by the smallest standard deviation of the bending angle difference $\alpha_{\text{obs}} - \alpha_{\text{clim}}$ found over a scale height (here, 7.5 kilometers) in the interval 60 to 80 kilometers [RD.18]. The noise σ_α is found by sliding a 7.5-kilometer wide window from 60 to 80 kilometers, computing the standard deviation within the window (Eq. I-1), and selecting the smallest value found:

$$\sigma_\alpha = \min \left\{ \frac{1}{n-1} \sum_{k=1}^n (\alpha_{\text{obs},k} - \alpha_{\text{clim},k})^2 \right\} \quad \text{I.1}$$

Here, n is the number of data points within the sliding window (for a 100 meter vertical grid spacing this number is 75), α_{obs} is the observed bending angle, and α_{clim} is the corresponding bending angles from the BAROCLIM (or, alternatively, MSIS) climatology.

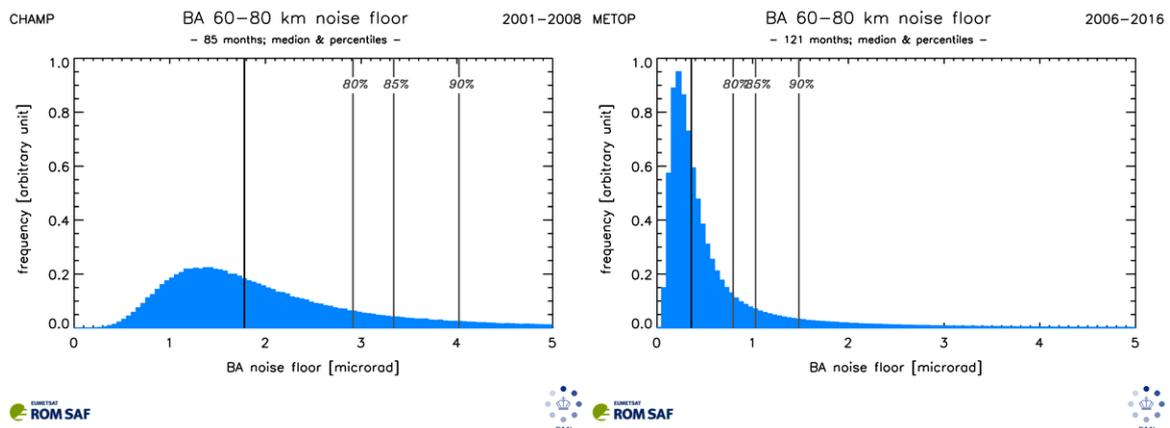


Figure I-1. Histograms showing the distributions of bending-angle standard deviations relative to a fitted background climatology. Data are from CHAMP (left panel) and Metop (right panel). The medians and the 80%, 85%, and 90% percentiles are indicated in the plots.

The resulting noise depends on the details of the processing from raw measured data to bending angles, particularly the effective degree of smoothing and the impact parameter grid used. Noise levels for different instruments, or for different time periods with the same instrument, can only be compared if the processing up to bending angle is the same.

I.2 Errors in observed refractivity and in the 1D-Var *a priori*

In one-dimensional variational assimilation, the following expression is minimized with respect to the atmospheric state \mathbf{x}

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{O}^{-1}(\mathbf{y}_o - H(\mathbf{x})) \quad \text{I.2}$$

Here, \mathbf{x}_b is a background state, or *a priori*, \mathbf{y}_o is the observation vector and H is the forward model. Matrices \mathbf{B} and \mathbf{O} describe the error covariances of the background and the observations (including any forward-modelling errors), respectively. Below, we also refer to a matrix \mathbf{H} , the Jacobian of H .

The state \mathbf{x}_s (the “solution”) that minimizes the above expression is statistically optimal in a maximum-likelihood sense if the observation and background errors are unbiased and Gaussian. Furthermore, the solution \mathbf{x}_s is a valid estimate of the true atmospheric state \mathbf{x} only if the error covariance matrixes used in the above expression are accurate descriptions of the true errors. Unfortunately, the error characteristics are not perfectly known.

Information on the errors can be gained from the statistics of the differences between observation, background, and solution – most conveniently expressed in observation space:

$$\mathbf{d}_{o-b} = \mathbf{y}_o - H(\mathbf{x}_b) \quad \text{I.3}$$

$$\mathbf{d}_{o-s} = \mathbf{y}_o - H(\mathbf{x}_s) \quad \text{I.4}$$

$$\mathbf{d}_{s-b} = H(\mathbf{x}_s) - H(\mathbf{x}_b) \quad \text{I.5}$$

This idea has been exploited by *Desroziers et al.* [2005] who developed a set of consistency criterias for the errors. If the background and observation errors are unbiased, Gaussian, and accurately describe the true errors, the following set of relations should hold:

$$E[\mathbf{d}_{s-b} \mathbf{d}_{o-b}^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T$$

$$E[\mathbf{d}_{o-s} \mathbf{d}_{o-b}^T] = \mathbf{O} \quad \text{I.7}$$

$$E[\mathbf{d}_{o-b} \mathbf{d}_{o-b}^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{O} \quad \text{I.8}$$

$$E[\mathbf{d}_{s-b} \mathbf{d}_{o-s}^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{O})^{-1} \mathbf{O} \equiv \mathbf{H}\mathbf{S}\mathbf{H}^T \quad \text{I.9}$$

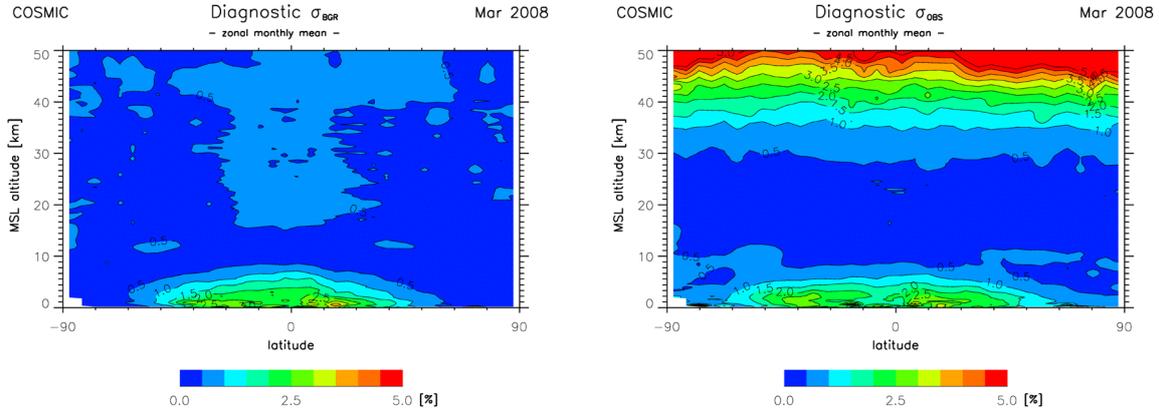


Figure I-2. Diagnosed errors given by equations I.12 and I.13. Data are from the RO instrument onboard the COSMIC-FM4 satellite in March 2008.

On the left hand side are the *diagnosed error covariances* based on the mean values of the differences in Eqs. I.3-5, while on the right hand side we find the error covariances assumed in the 1D-Var retrieval in Eq. I.2. The matrix \mathbf{HBH}^T describes the covariances of background errors linearly transformed to observation space. Similarly, the matrix \mathbf{HSH}^T describes the covariances of solution errors (\mathbf{S}) linearly mapped to observation space.

The diagnosed errors in Eqs. I.6-9 provide a means to *monitor the stability of errors*. If, during a long time period, the assumed error covariances used in the 1D-Var retrieval are kept unchanged, then any significant changes of the diagnosed error covariances during that time period must be due to changes of the actual errors in the background or in the observations.

In the ROM SAF climate monitoring, the mean diagonal elements of the diagnosed error covariances in Eqs. I.6 and I.7 are continuously monitored month by month:

$$\langle \sigma_{b,i}^2 \rangle = \frac{1}{M_i} \sum_{j=1}^{M_i} (y_{s,j} - y_{b,j})(y_{o,j} - y_{b,j}) \quad \text{I.10}$$

$$\langle \sigma_{o,i}^2 \rangle = \frac{1}{M_i} \sum_{j=1}^{M_i} (y_{o,j} - y_{s,j})(y_{o,j} - y_{b,j}) \quad \text{I.11}$$

Here, the averaging includes all data at a certain height within a latitude band and time interval (in the ROM SAF climate monitoring the bin sizes are 5 degrees of latitude and one calendar month). Index i denotes a latitude band, and index j loops over the M_i observations in latitude band i . Note that $\sigma_{o,i}$ and $\sigma_{b,i}$ are vertical profiles, i.e. they are functions of height.

Since the observed quantity is refractivity, which falls off exponentially with height, the diagnosed errors are more conveniently expressed in relative terms. Thus, the quantities that are actually monitored, and for which zonal plots are produced, are:

$$\sigma_{b,i}^{\text{diag}} = 100 \cdot \sqrt{\langle \sigma_{b,i}^2 \rangle} / \langle y_{b,i} \rangle \quad \text{I.12}$$

$$\sigma_{o,i}^{\text{diag}} = 100 \cdot \sqrt{\langle \sigma_{o,i}^2 \rangle} / \langle y_{o,i} \rangle \quad \text{I.13}$$

where, once again, the averages include all observations in a latitude band and a calendar month. An example of the monitoring quantities for data from the COSMIC-FM4 satellite from March 2008 is shown in Figure I-2.