EUMETSAT

ROM SAF

RADIO OCCULTATION METEOROLOGY

Algorithm Theoretical Baseline Document: Level 2B and 2C 1D-Var products

Version 4.3

15 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)

Ref: SAF/ROM/DMI/ALG/1DVAR/002

DOCUMENT AUTHOR TABLE

	Author(s)	Function	Date
Prepared by:	Johannes K. Nielsen	ROM SAF Scientist	15/10/21
Reviewed by (Internal):	Stig Syndergaard	ROM SAF Design Manager	21/02/20
Approved by:	Kent B. Lauritsen	ROM SAF Project Manager	15/10/21

DOCUMENT CHANGE RECORD

Version	Date	By	Description
0.9	28/02/08	HGL	Draft document.
1.0	12/03/08	HGL	Release version. KBL comments.
2.0	11/05/09	KMK	Version for ORR-B. Since v1.0 the GRAS 1DV software is modified to rely much more heavily on ROPP 2.0. Reliance on OPTESTRET code removed.
2.1	30/07/09	KMK	Updated (non-released) version based on ORR-B comments: Background errors modified and background data source modified (3 hour time steps).
2.2	03/07/2012	JKN	Including editorial remarks from ORR-B part 1 close-out; converted doc to LaTeX; changed background error covariance calculation to new method based on averaging “error of first guess” (ef)-fields. Comments provided by Sean Healy included. A draft version of this document was sent to SG for comments.
2.3	06/05/2014	JKN	Version submitted for ORR4 and ORR-B-Backlog review. Included comments from EUMETSAT (November 2012) and ECMWF (February 2014); New figures from test run of 1DV 2.6.2. Updated background error covariances derived from ”ses” error fields.
2.4	23/06/2014	JKN	Updated for ORR4 ORR B Backlog review; RID implemented: 001, 002, 003, 004, 005, 006, 007, 008, 009, 012, 028, 031, 036, 039
2.5	08/03/2016	JKN	Version submitted for the PCR-RE1 review. Section 6.3 has been added specifically for the reprocessing. Other parts of the document were restructured and extended: Chap 3 (former Chap 2) has been pruned to document only the parts that are common to NRT, offline and reprocessing. Former Chap. 4 has moved to 4.1. Former 2.6 has been moved to to 4.3, while sections 4.2, 4.4-5 have been added. Sections 6.1 and 6.2 contains the parts of former Chap 2, specific to NRT and offline. Smaller adjustment has been made throughout the document to smooth transitions between sections and to obey the new common ROM SAF ATBD template.

DOCUMENT CHANGE RECORD

... continued

Version	Date	By	Description
2.6	18/04/2016	JKN	Updated version for the PCR-RE1 review. Implemented RID 2, 15, 16, 31, 32 and 33. Inserted Desrozier plot to justify observation error assumption.
2.7	09/05/2016	JKN	Implemented further adjustments according to RID's from PCR-RE1 review. RID 3, 15, 16,31, 32 and 33. Fig. 6.4 is updated.
2.8	22/01/2018	JKN	This version is an internal version. List of changes: - Ch. 1,3,4. Updates related to 1DV v.3.3. - Appendices I,II,III. Explanation of change in 1DV configuration: (-logq and -logp) switched off. - Ch. 3,4,6 Description of new B matrix for RE1; latitude resolved uncertainty, inflated surface pressure uncertainty. - Appendix I and IV. Description of changed method for full level pressure calculation according to Simmons et. al 1981. - Ch. 1 to Appendix IV. General revision of entire document with improved language. - Ch. 6,Appendix I. Removed irrelevant comments related to old 8 km restriction. - Ch. 4,6. Produced new figures as consequence new background, analysis and background error implementations (figure 4.1, 4.3, 6.7, 6.8, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17, 6.18), - Ch. 6. Swapped sections 6.2 and 6.3. - Ch. 6. Completely rewrote section about contamination of ERA-I background, (section 6.2.4).
2.9	10/06/2018	JKN	Version prepared for DRR-RE1 review. List of changes: - Ch. 6. Clarified use of 2012 ses (91 L). - misprints caught by S. Healy corrected. - front page format change - Page 1-3 format change - headers, eq. numbers, changed - editorial changes - Annex 4. Figure captions expanded
3.0	02/09/2018	JKN	Updated version implementing the following RIDS for DRR-RE1 & ORRs review: - RIDs 119, 127, 130, 131, 132, 133, 134, 135, 171, 172, 173, 174, 176, 177, 178, 179, 180, 181, 182, 183, 184,185, 186, 187, 188, 189, 190, 191, 192, 193: Editorial changes implemented. - RID 175: Ch. 3, "error" exchanged with "uncertainty" where possible.
3.1	20/12/2018	JKN	Updated version based on ICDR concept discussions at ROM SAF SG22: - p5 CDR/ICDR descriptions inserted. - Sec 1.1 New GRM numbers for ICDR and Metop C - p16 Definitions to incl. CDR/ICDR - Sec 3.2 CDR/ICDR clarification - Sec 3.7 CDR/ICDR clarification - Sec 4.3 footnote CDR/ICDR clarification - Ch 6 CDR/ICDR clarification

DOCUMENT CHANGE RECORD *... continued*

Version	Date	By	Description
4.0	21/02/2020	JKN	Updated version for EPS-SG PDCR review: - Sec 1.1 updated version numbers - Sec 1.1.1 inserted EPS-SG description - Tab 1.2 inserted extra tab EPS-SG products - p 12 EPS-SG in acronyms & abbbrew - p0-14 Aligned preamble with current template10.tex - p14 ERA5 in acronyms - p21 Removed specific Metop A/B reference input data list includes ERA5 - p21 mention ERA5 in footnote (1) - p26 removed “so called” - p32 changed output description to match 1DV-v4.2 - p32 changed table layout - p36 removed version numbers in footnote - p37 inserted footnote about $q \geq 10^{-6}$ - p39 generalized references to missions and code versions. - p43 inserted description of 60 L ERA5 proc. - p46 mentioning GPAC numbers - p51 transition to ERA5 in offline described - p53 updated 1DV version table - p55 appendix II was updated to reflect the current 1DV version (4.2) - p56 inserted new configuration file (v4.2) - p64 removed misleading sentence about constant “C” at top level Updated version for ORR12 review: - None
4.1	14/05/2020	JKN	Updated version prepared for ORR12 Closeout implementing the following: - Sec. 6.2.2: last par. Use of Simmons et al [RID 033, 034] - p. 23: footnote: Removed 137 levels [RID 023] - App. IV: Future use of Simmons et al [RID 033, 034] - Sec. 6.3: explained tropopause fallback [RID 024] - Sec. 5.2: footnote: Explained bug. [RID 028] - Sec. 6.3 and 5.2: offline prod. unstable [RID 32] - Sec. 3.7.2: clarified number of levels [RID 046] - Sec. 5.2: removed “constraints” from title - Sec. 1.1: text about future EPS-SG Day 1 products removed [RIDs 43, 98] - Sec. 1.1.1: text and table about future EPS-SG Day 1products removed [RIDs 43, 98] - Sec. 1.3: removed EPS-SG from acronyms [RIDs 43, 98]

DOCUMENT CHANGE RECORD *... continued*

Version	Date	By	Description
4.2	28/09/2021	JKN	Updated version submitted for the PCR review for Sentinel-6 NTC products implementing the following changes: - Sec. 1.1: New GRM numbers for Sentinel 6 NTC - Sec. 1.4: NTC mentioned in definitions - Sec. 3.2: mentioning NTC in table 3.1 - p. 31: NTC mentioned in footnote - p. 37: NTC (q<0 fix) mentioned in footnote - Chapter 6, p 39: NTC description inserted. - Sec. 6.3 Description of NTC processing
4.3	15/10/2021	JKN	Updated version for the PCR review for Sentinel-6 NTC products implementing the following changes: - p. 14,40,59: Sentinel 6 -> Sentinel-6 - Sec. 1.4 “Sentinel-6”, “about”, “applicable” inserted - p. 18: NTC in acronyms - p. 59: ITC->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 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

Document Change Record 2

List of Contents 7

1 Introduction 8
1.1 Purpose of the Document . . . 8
1.2 Applicable & Reference documents . . . 15
1.2.1 Applicable documents . . . 15
1.2.2 Reference Documents . . . 15
1.3 Acronyms and abbreviations . . . 18
1.4 Definitions . . . 19
1.5 Document overview . . . 20

2 Algorithm overview 21

3 Algorithm description 22
3.1 From Refractivity to Pressure, Temperature, and Humidity . . . 22
3.2 1DV input data . . . 23
3.3 The 1D-Var Level Structure . . . 24
3.4 The Forward Model . . . 25
3.5 The 1D-Var Cost Function . . . 25
3.6 Minimising the Cost Function . . . 26
3.7 Observational/Background Error Covariances . . . 27
3.7.1 Observational error covariance . . . 27
3.7.2 Background error covariance structure . . . 28
3.8 Reduction of Background Information Error . . . 30

4 Practical Considerations 32
4.1 Evaluating the Retrieved Solution . . . 32
4.2 Validating the Retrieved Solution . . . 32
4.3 Quality control . . . 34
4.4 Exception handling . . . 35
4.5 Outputs . . . 35
4.5.1 Example of retrieved meteorological fields . . . 38

5 Assumptions and Limitations 40
5.1 Assumptions . . . 40
5.2 Limitations . . . 40

6 Description of differences for NRT, Offline, NTC, CDR1 and ICDR1 42
6.1 Configuration for ROM SAF 1D-Var, NRT products . . . 42
6.1.1 Observations from Metop 1st and 2nd generation . . . 42
6.1.2 Background profiles from ECMWF Forecasts . . . 42
6.1.3 Mean background uncertainty estimates from Scaled Ensemble Standard Deviation . . . 43
6.2 Configuration for first ROM SAF reprocessing, RE1 (1DV-3.3 and 1DV-3.4) 47
6.2.1 Observations from multiple satellite missions. . . . 47

6.2.2 Background from ECMWF ERA Interim and ERA5 47
6.2.3 Mean background uncertainty estimates from ERA-I error of first guess estimates 49
6.2.4 Special properties of the RE1 background data. 51
6.3 Configuration for Offline processing (1DV-3.3 and 1DV-4.2) 57
6.4 Configuration for NTC processing (1DV-4.2.2) 57

Appendices **58**
Appendix I. Intermediate changes to the 1D-Var code and configuration since version 1DV version 3.0 58
Appendix II. ROM SAF 1DV Software 60
Appendix III. The configuration file `romsaf_ecmwf_refrac_1dvar.cf(v4.2)` 61
Appendix IV. Correction to full level pressure interpolation. 66

1 Introduction

1.1 Purpose of the Document

This ATBD document describes the algorithms used to derive the Level 2B and 2C 1D-Var products produced by the Radio Occultation Meteorology (ROM) Satellite Application Facility (SAF). The document covers NRT, Offline, CDR and ICDR-products. The complete list of products covered by this ATBD is provided in Table 1.1. Note that this table may include both products in development and 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 Product Requirement Document (PRD) [AD.3]. The ATBD software package is denoted "1DV" and is based on ROPP[1] routines. The current 1DV version, 1DV-v4.2, is implemented for production of ROM SAF Offline-v1.1 and ROM SAF ICDR-v1.1. It is based on ROPP 9.0 with adaptations made by DMI. Adaptations made by DMI will be included in a future official ROPP release. The NRT production, is still running on an older version; 1DV-v3.0. The ROM SAF CDR-v1.0 was produced with 1DV-v3.3 See Appendix I for details on various 1DV-versions.

Table 1.1: *List of products covered by this ATBD:*
**) The fifth column describes the original data source; satellite/instrument, data-level, institution. "Level 1A" refers to excess phase data. See also [RD.26]*

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
NRT						
GRM-02	NRT Temperature Profile	NTPMEA	NRT	Metop-A/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-03	NRT Specific Humidity Profile	NHPMEA	NRT	Metop-A/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-04	NRT Pressure Profile	NPPMEA	NRT	Metop-A/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-05	NRT Surface Pressure	NSPMEA	NRT	Metop-A/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-41	NRT Temperature Profile	NTPMEB	NRT	Metop-B/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-42	NRT Specific Humidity Profile	NHPMEB	NRT	Metop-B/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-43	NRT Pressure Profile	NPPMEB	NRT	Metop-B/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-44	NRT Surface Pressure	NSPMEB	NRT	Metop-B/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF

Continued on next page ...

[1] ROPP stands for "Radio Occultation Processing Package". It is a software package maintained by ROM SAF. See [RD.20] for an ROPP overview

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
GRM-61	NRT Temperature profile	NTPMEC	NRT	Metop-C/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-62	NRT Specific Humidity profile	NHPMEC	NRT	Metop-C/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-63	NRT Pressure Profile	NPPMEC	NRT	Metop-C/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF
GRM-64	NRT Surface Pressure	NSPMEC	NRT	Metop-C/ GRAS	GTS EUMETCast Web	BUFR BUFR/netCDF BUFR/netCDF

Continued on next page ...

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
OFL						
GRM-10	OFL Temperature Profile	OTPMEA	OFL	Metop-A/ GRAS	Web	BUFR netCDF
GRM-11	OFL Specific Humidity Profile	OHPMEA	OFL	Metop-A/ GRAS	Web	BUFR netCDF
GRM-12	OFL Pressure Profile	OPPMEA	OFL	Metop-A/ GRAS	Web	BUFR netCDF
GRM-13	OFL Surface Pressure	OSPMEA	OFL	Metop-A/ GRAS	Web	BUFR netCDF
GRM-48	OFL Temperature Profile	OTPMEB	OFL	Metop-B/ GRAS	Web	BUFR netCDF
GRM-49	OFL Specific Humidity Profile	OHPMEB	OFL	Metop-B/ GRAS	Web	BUFR netCDF
GRM-50	OFL Pressure Profile	OPPMEB	OFL	Metop-B/ GRAS	Web	BUFR netCDF
GRM-51	OFL Surface Pressure	OSPMEB	OFL	Metop-B/ GRAS	Web	BUFR netCDF

Continued on next page ...

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
GRM-61	OFL Temperature profile	OTPMEC	OFL	Metop-C/ GRAS	Web	BUFR netCDF
GRM-62	OFL Specific humidity profile	OHPMEC	OFL	Metop-C/ GRAS	Web	BUFR netCDF
GRM-63	OFL Pressure Profile	OPPMEC	OFL	Metop-C/ GRAS	Web	BUFR netCDF
GRM-64	OFL Surface Pressure	OSPMEC	OFL	Metop-C/ GRAS	Web	BUFR netCDF

Continued on next page ...

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
REP						
GRM-29-L2-T-R1	Reprocessed Temperature Profile	RTPMET	CDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-H-R1	Reprocessed Specific Humidity Profile	RHPMET	CDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-P-R1	Reprocessed Pressure Profile	RPPMET	CDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-S-R1	Reprocessed Surface Pressure	RSPMET	CDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-30-L2-T-R1	Reprocessed Temperature Profile	RTPCO1	CDR	COSMIC Level 1A from CDAAC	Web	BUFR netCDF
GRM-30-L2-H-R1	Reprocessed Specific Humidity Profile	RHPCO1	CDR	COSMIC Level 1A from CDAAC	Web	BUFR netCDF
GRM-30-L2-P-R1	Reprocessed Pressure Profile	RPPCO1	CDR	COSMIC Level 1A from CDAAC	Web	BUFR netCDF
GRM-30-L2-S-R1	Reprocessed Surface Pressure	RSPCO1	CDR	COSMIC Level 1A from CDAAC	Web	BUFR netCDF
GRM-32-L2-T-R1	Reprocessed Temperature Profile	RTPCHA	CDR	CHAMP Level 1A from CDAAC	Web	BUFR netCDF
GRM-32-L2-H-R1	Reprocessed Specific Humidity Profile	RHPCHA	CDR	CHAMP Level 1A from CDAAC	Web	BUFR netCDF
GRM-32-L2-P-R1	Reprocessed Pressure Profile	RPPCHA	CDR	CHAMP Level 1A from CDAAC	Web	BUFR netCDF
GRM-32-L2-S-R1	Reprocessed Surface Pressure	RSPCHA	CDR	CHAMP Level 1A from CDAAC	Web	BUFR netCDF
GRM-33-L2-T-R1	Reprocessed Temperature Profile	RTPGHA	CDR	GRACE Level 1A from CDAAC	Web	BUFR netCDF
GRM-33-L2-H-R1	Reprocessed Specific Humidity Profile	RHPGHA	CDR	GRACE Level 1A from CDAAC	Web	BUFR netCDF
GRM-33-L2-P-R1	Reprocessed Pressure Profile	RPPGHA	CDR	GRACE Level 1A from CDAAC	Web	BUFR netCDF
GRM-33-L2-S-R1	Reprocessed Surface Pressure	RSPGHA	CDR	GRACE Level 1A from CDAAC	Web	BUFR netCDF

Continued on next page ...

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
GRM-29-L2-T-I1	Reprocessed Temperature Profile	ITPMET	ICDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-H-I1	Reprocessed Specific Humidity Profile	IHPMET	ICDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-P-I1	Reprocessed Pressure Profile	IPPMET	ICDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF
GRM-29-L2-S-I1	Reprocessed Surface Pressure	ISPMET	ICDR	Metop Level 1A data from EUM Secretariat	Web	BUFR netCDF

Continued on next page ...

Product ID	Product name	Product acronym	Product type	Operational satellite input *	Dissemination means	Dissemination format
NTC						
GRM-118	NTC Temperature Profile	OTPS6	NTC	Sentinel-6 Level 1B data from EUM Secretariat	Web	BUFR netCDF
GRM-119	NTC Specific Humidity Profile	OHPS6	NTC	Sentinel-6 Level 1B data from EUM Secretariat	Web	BUFR netCDF
GRM-120	NTC Pressure Profile	OPPS6	NTC	Sentinel-6 Level 1B data from EUM Secretariat	Web	BUFR netCDF
GRM-121	NTC Surface Pressure	OSPS6	NTC	Sentinel-6 Level 1B data from EUM Secretariat	Web	BUFR netCDF

1.2 Applicable & 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 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, Ref. 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, The Radio Occultation Processing Package (ROPP): ROPP I/O Reference Manual, Ref: SAF/GRAS/METO/RM/ROPP/002, -.

[RD.2] ROM SAF, The Radio Occultation Processing Package (ROPP): ROPP Forward Model Reference Manual, Ref: SAF/GRAS/METO/RM/ROPP/003, -.

[RD.3] ROM SAF, The Radio Occultation Processing Package (ROPP): ROPP UTILS Reference Manual, Ref: SAF/GRAS/METO/RM/ROPP/001, 2008.

[RD.4] Aparicio, J. M. and Laroche, S., An evaluation of the expression of the atmospheric refractivity for GPS signals, *Journal of Geophysical Research (Atmospheres)*, *116*, 11 104, 2011.

[RD.5] Bevis, M., Businger, S., Chiswell, S., Herring, T. A., Anthes, R. A., Rocken, C., and Ware, R. H., Mapping zenith wet delays onto precipitable water, *J. Appl. Meteor.*, *33*, 379–386, 1994.

[RD.6] Desroziers, G., Berre, L., Chapnik, B., and Poli, P., Diagnosis of observation, background and analysis-error statistics in observation space, *Quarterly Journal of the Royal Meteorological Society*, *131*, 3385–3396, URL `http://dx.doi.org/10.1256/qj.05.108`, 2005.

[RD.7] ECMWF, http://www.ecmwf.int/products/data/technical/model_levels/index.html and http://www.ecmwf.int/... research/ifsdocs/DYNAMICS/Chap2_Discretization4.html#961180, 2018.

[RD.8] Eyre, J. R., Kelly, G. A., McNally, A. P., Andersson, E., and Persson, A., Assimila-

tion of TOVS radiance information through one–dimensional variational analysis, *Quart. J. Roy. Meteorol. Soc.*, *119*, 1427–1463, 1993.

[RD.9] Healy, S. B. 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.10] Holm, E. V. and Kral, T., Flow-dependent, geographically varying background error covariances for 1D-VAR applications in MTG-IRS L2 Processing, ECMWF Technical Memorandum No. 680, 2012.

[RD.11] Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R., Observing earth's atmosphere with radio occultation measurements using the Global Positioning System, *J. Geophys. Res.*, *102*, 23.429–23.465, 1997.

[RD.12] Kursinski, E. R., Healy, S. B., and Romans, L. J., Initial results of combining GPS occultations with ECMWF global analyses within a 1DVar framework, *Earth Planets Space*, *52*, 885–892, 2000.

[RD.13] Lorenc, A. C., Analysis methods for numerical weather prediction, *Quart. J. Roy. Meteorol. Soc.*, *112*, 1177–1194, 1986.

[RD.14] Press, W., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical recipes in Fortran – The Art of Scientific Computing*, Cambridge University Press, Cambridge, New York, 2nd edn., 1992.

[RD.15] Randel, W. J., Wu, F., and Gaffen, D. J., Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalysis, *J. Geophys. Res.*, *105*, 15.509–15.523, 2000.

[RD.16] Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote sounding measurements of thermal radiation, *Rev. Geophys. Space Phys.*, *14*, 609–624, 1976.

[RD.17] Rodgers, C. D., *Inverse methods for atmospheric sounding: Theory and practice*, World Scientific Publishing, Singapore, New Jersey, London, Hong Kong, 2000.

[RD.18] ROM SAF, The Radio Occultation Processing Package (ROPP): ROPP 1D-Var Reference Manual, Ref: SAF/ROM/METO/RM/ROPP/004, -.

[RD.19] ROM SAF, The Radio Occultation Processing Package (ROPP): 1D-Var,User Guide, Ref: SAF/GRAS/METO/UG/ROPP/003, -.

[RD.20] ROM SAF, The Radio Occultation Processing Package (ROPP) Overview, SAF/ROM/METO/UG/ROPP/001, -.

[RD.21] ROM SAF, Validation report: GRM-02, 03, 04, 05, SAF/ROM/DMI/RQ/REP/002, .

[RD.22] ROM SAF, Algorithm Theoretical Baseline Document: Level 3 Gridded Data, SAF/ROM/DMI/ALG/GRD/001, .

[RD.23] ROM SAF, Report 01: Mono-dimensional data thinning for GPS Radio Occultations, SAF/GRAS/METO/REP/GSR/001, 2007.

[RD.24] ROM SAF, Report: Levenberg-Marquardt minimisation in ROPP, SAF/GRAS-/METO/REP/GSR/006, 2008.

[RD.25] ROM SAF, Report: Refractivity coefficients used in the assimilation of GPS radio occultation measurements, SAF/GRAS/METO/REP/GSR/009, 2009.

[RD.26] ROM SAF, Algorithm Theoretical Baseline Document: Level 1B Bending angles., Ref. SAF/ROM/DMI/ALG/BA/001, Version 2.0, 2020.

[RD.27] ROM SAF, Algorithm Theoretical Baseline Document: Level 2A Refractivity profiles., Ref. SAF/ROM/DMI/ALG/REF/001, Version 2.0, 2020.

[RD.28] Rueger, J. M., Refractive index formulae for electronic distance measurement with radio and millimetre waves, Unisurv Report S-68. School of Surveying and Spatial Information Systems, University of New South Wales, [Summary at http://www.fig.net/pub/fig_2002/Js28/JS28_rueger.pdf], 2002.

[RD.29] Scherllin-Pirscher, B., Steiner, A. K., Kirchengast, G., Kuo, Y.-H., and Foelsche, U., Empirical analysis and modeling of errors of atmospheric profiles from gps radio occultation, *Atmospheric Measurement Techniques*, *4*, 1875–1890, URL `http://www.atmos-meas-tech.net/4/1875/2011/`, 2011.

[RD.30] Simmons, A. J. and Burridge, D. M., An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates, *Mon. Wea. Rev.*, *109*, 758–766, 1981.

[RD.31] Smith, E. K. and Weintraub, S., The constants in the equation for atmospheric refractivity index at radio frequencies, in *Proc. IRE*, vol. 41, pp. 1035–1037, 1953.

[RD.32] Tarantola, A. T. and Valette, B., Generalized non-linear inverse problems solved using the least squares criterion, *Reviews of Geophysics and Space Physics*, *20*, 219–232, 1982.

[RD.33] Yang, S. and Zou, X., Assessments of cloud liquid water contributions to GPS radio occultation refractivity using measurements from COSMIC and CloudSat, *Journal of Geophysical Research-Atmospheres*, *117*, D06 219, 2012.

[RD.34] Zou, X., Yang, S., and Ray, P. S., Impacts of ice clouds on GPS radio occultation measurements, *Journal of the Atmospheric Sciences*, *69*, 3670–3682, 2012.

1.3 Acronyms and abbreviations

BUFR	Binary Universal Form for the Representation of meteorological data
CDR	Climate Data Record
CHAMP	Challenging Mini–Satellite Payload
COSMIC	Constellation Observing System for Meteorology, Ionosphere & Climate
DMI	Danish Meteorological Institute
ECMWF (EUMETSAT)	European Centre for Medium-Range Weather Forecasts
ERA-I	ERA-Interim, ECMWF reanalysis
ERA5	Fifth major global reanalysis produced by ECMWF
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GNSS	Global Navigation Satellite Systems (generic name for GPS, GLONASS, GALILEO, etc.)
GPS	Global Positioning System (US)
GRACE	Gravity Recovery and Climate Experiment.
GRAS	GNSS Receiver for Atmospheric Sounding (onboard Metop)
GTS	Global Telecommunications System (The core of WIS)
IFS	Integrated Forecasting System
ICDR	Interim Climate Data Record
IWC	Ice Water Content
LEO	Low Earth Orbit
Metop	Meteorological operational polar satellites (EUMETSAT)
NTC	Not Time Critical
netCDF	network Common Data Form (Unidata)
NRT	Near Real Time
NWP	Numerical Weather Prediction
OFL	Offline
PRD	Product Requirements Document
RE1	1st ROM SAF reprocessing
RO	Radio Occultation
ROM	Radio Occultation Meteorology
ROPP	Radio Occultation Processing Package
SAF	Satellite Application Facility (EUMETSAT)
STDV	Standard Deviation
VAR	Variational analysis; 1D, 2D, 3D or 4D versions (NWP data assimilation technique)
WIS	WMO Information System
1DV	The software used for ROM SAF operational 1D-Var processing (based on ROPP)

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 and product types are defined below[2]. 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 software[3]. 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;

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

[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`

latency than the system used to generate the CDR[4].

1.5 Document overview

The contents of the document are as follows:

Chapter 2 contains a brief algorithm overview.

Chapter 3 contains description of the parts of the algorithm which are common to all products.

Chapter 4 contains practical considerations.

Chapter 5 describes assumptions and limitations.

Chapter 6 contains details that are specific to the different data types NRT, Offline, NTC, CDR and ICDR.

Finally appendix I lists intermediate changes to 1DV code and configuration for intermediate updates since last release, appendix II describes the software, appendix III is a print of the common ROPP configuration file used for all 1DV products and appendix IV describes the effects of changing full level pressure interpolation methods.

[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 Algorithm overview

The GNSS Radio Occultation Instrument onboard a low earth orbit (LEO) satellite such as Metop measures the phase shift of the radio signal from a GNSS satellite as it sets or rises behind the Earth's limb as seen from the LEO. This phase shift as a function of time is converted to bending angle and subsequently, through the Abel transform, to refractivity as a function of height at a particular location in the Earth's atmosphere [RD.27]. The refractivty is a function of both dry air density and the partial pressure of water vapor. In the troposphere water vapour contributes to the refractivity. This causes an ambiguity in the interpretation of the radio occultation profile in the sense that it cannot be known to which extent the refractivity signal is influenced by water vapour. Therefore the conversion of refractivity or bending angle to temperature, pressure and humidity requires some additional information, e.g. from an atmospheric model. The one-dimensional variational (1D-Var) technique provides an established method to estimate this meteorological information from the more fundamental refractivity or bending angle data.

The 1DV software used for operational 1D-Var processing within the ROM SAF is based on the ROPP code developed within the ROM SAF project. The ROPP package is available from the web site http://www.romsaf.org.

The input to the 1D-Var processing consists of:

- refractivity as a function of geopotential height, together with latitude, longitude and time of the occultation;
- a background profile of pressure, temperature, humidity, surface pressure and surface geopotential height interpolated to the observed latitude and longitude;
- an estimate of the error covariances of the retrieved refractivity profile;
- an estimate of the error covariances of the background profile;

The output from the 1D-Var processing consists of:

- the most likely vertical profile of temperature, humidity, and surface pressure that is consistent with both the observations and the background to within their respective errors;
- theoretical uncertainty, i.e. a solution error covariance matrix for temperature, humidity, and surface pressure;

The input and output data are stored in netCDF files following the ROPP definitions tailored for radio occultation data [RD.1]. The output data is also stored in BUFR format.

The uncertainties of atmospheric variables (refracticity, temperature, specific humidity and pressure) represent the expected deviations from their true values expressed in standard deviations and biases. The derived uncertainties that come with each product are calculated from the uncertainties of the input variables (measurement and background uncertainties).

3 Algorithm description

This section gives a general overview of 1D-Var processing of RO data.

3.1 From Refractivity to Pressure, Temperature, and Humidity

The refractivity (N) is related to the temperature (T) and the partial pressures of dry air (P_d) and water vapour (e) through:

$$N = \kappa_1 \frac{P_d}{Z_d T} + \kappa_2 \frac{e}{Z_w T^2} + \kappa_3 \frac{e}{Z_w T} \tag{3.1}$$

where $\kappa_1 = 77.643$ K/hPa, $\kappa_2 = 3.75463 \cdot 10^5$ K^2/hPa and $\kappa_3 = 71.2952$ K/hPa are empirically determined constants, and Z_d and Z_w are the dry air and wet air compressibility factors [RD.5, RD.9, RD.11, RD.31, RD.28, RD.25, RD.4]. The compressibility factors Z_* correct for non-ideal gas effects. Z_d and Z_w may be up to 0.05% smaller than unity in the denser parts of the atmosphere [RD.25].

The recommendations of [RD.28] and [RD.25] are followed. That is, the functional form of the refractivity and the coefficients are adopted from these references. Here it is assumed that scattering due to ionization is negligible and the impact of precipitation on the refractivity is neglected. In [RD.33] it is estimated that water droplets can contribute with a refractivity bias as high as 1.2 % for some clouds. The contribution to refractivity from ice amounts to 0.6% of the total refractivity, at 1 g/m^3 IWC[RD.34].

In a dry atmosphere, assuming hydrostatic equilibrium, the gas law and a reference pressure at a single altitude, one can completely determine the vertical profile of pressure, temperature, and density from the retrieved refractivities. However, in the troposphere e may not be negligible, and as a consequence the inversion problem becomes ambiguous. It is impossible to tell the difference between effects caused by temperature and effects caused by humidity solely from a refractivity profile. To solve this ambiguity, some additional information on the atmospheric state is required.

One way to deal with this problem is to assume an a priori humidity profile and then derive the temperatures, or to assume an a priori temperature profile and then derive the humidities from the measured refractivities. However, the measurements and the a priori information are both affected by uncertainties and the results can be difficult to interpret.

The 1D-Var technique provides an established method to combine the mutual information provided by observations and some a priori data taken from an atmospheric model [RD.8, RD.9, RD.12]. The algorithm attempts to minimize a cost function that takes into account both the background data, the observation data and their respective uncertainties and error correlations (see section 3.5). As a result, one gets statistically optimal pressure, temperature, and humidity profiles consistent with both the RO measurements and the background model and their respective errors. The general approach to variational retrieval has been described in detail by many authors [RD.13, RD.16, RD.32].

3.2 1DV input data

The ROM SAF 1D-Var utilizes various sources of input data which are listed in table 3.1. Even though the observations come from various satellite missions, specific for each product, they are all processed to Level 2A (refractivity) within the ROM SAF. The background data are retrieved from various ECMWF data sets, all generated by different versions of the IFS. Therefore the overall format is the same in all background data sets, but the number of levels as well as the error covariance estimates varies among different Level 2B and Level 2C products. The error covariance estimates are derived using error estimates and a fixed correlation matrix provided by ECMWF. The error estimates are disseminated by ECMWF daily as the "ses" (Scaled Ensemble Standard Deviation) fields, and in the ERA Interim reanalysis data as the "ef" (Error of First Guess) fields.

Table 3.1: *Table showing a list of all input data for the ROM SAF 1D-Var algorithm*

Input	Provider	Note
Refractivity observations	ROM SAF	Metop bending angles are disseminated by EUMETSAT and processed to refractivity at DMI. The reprocessed data sets also include refractivities based on the COSMIC, GRACE and CHAMP missions.
Background profiles	ECMWF	For NRT and NTC production; GRIB files (fields) are retrieved through the ECMWF DIS system with 6 hours intervals and atmospheric state vectors (profiles) are extracted at DMI for each occultation. Nearest forecast in time is used. For reprocessed and offline production; 3 hour interval forecasts from ERA-Interim and ERA5 are used with time interpolation between the two nearest forecasts.
Background error correlations	ECMWF	Constant matrix (produced in 2012).
Background error standard deviations	ECMWF	For NRT production: Constant error STDV for temperature and constant relative error for humidity and surface pressure. STDVs are computed from "ses" error estimates valid for 2012. For reprocessed (CDR and ICDR) Offline and NTC data: Latitude resolved error STDV for temperature, humidity and surface pressure. STDVs are computed from "ef" (ERA-Interim) error estimates throughout the RO era[i].

[i] From August 2019 the ICDR is based on background profiles from ERA5, but the background error model based onERA-I "ef" is continuously being used also after introduction of ERA5 background profiles.

3.3 The 1D-Var Level Structure

The 1D-Var calculation in the ROPP code can be performed using either of 2 model level definitions [RD.1]: The hybrid-sigma (ECMWF-model) definition or the geopotential-based (Met Office-model) definition. Since the 1DV implementation is based on background data from ECMWF forecasts, the calculation is done on hybrid-sigma (ECMWF) levels. In this

level definition temperature and humidity are specified on $\frac{n-1}{2}$ model levels [1][2][RD.7]. The pressures of these model levels are uniquely determined from the surface pressure. See also section 6.1.2.

During minimization of the cost function $J(\boldsymbol{x})$ in Eq. (3.3) the minimizer algorithm varies surface pressure, temperature, and humidity. Since the geopotential height of a given pressure level is a function of $\boldsymbol{T}$ and $\boldsymbol{q}$, the geopotential heights of the 1D-Var levels will generally change during the optimization process even if the surface pressure does not.

3.4 The Forward Model

The forward model $H(\boldsymbol{x})$ maps the state vector information $\boldsymbol{x}$ into measurement space. For the definitions in the 1DV software package, the observation vector, $\boldsymbol{y}^o$, is refractivity as function of geopotential height, while the state vector, $\boldsymbol{x}$, is a n-component vector containing surface pressure (in hPa) and, for each of the $(n-1)/2$ atmospheric levels, temperature (in K) and specific humidity (in g/kg).

The forward model computes the refractivities corresponding to the state $\boldsymbol{x}$, at the model levels, using Eq. (3.1). It then computes the geopotential heights of the model levels by integrating the hydrostatic equation, using the surface pressure and geopotential at the surface as boundary conditions. Finally the forward model interpolates the refractivity N to the geopotential heights of the observation vector.

The gradient matrix $\boldsymbol{H}(\boldsymbol{x})$ i.e. the derivative of H with respect to the atmospheric state $\boldsymbol{x}$ is used to find the minimum of the cost function $J(\boldsymbol{x})$, as seen in Section 4.1. It is also used to map the background errors $\boldsymbol{B}$ into observation space through $\boldsymbol{HBH}^T$, and to estimate the errors associated with the 1D-Var solution. For a linear forward model model, the theoretically estimated solution error covariance matrix $\boldsymbol{S}$ (sometimes called $\boldsymbol{A}$)[3] is given by

$$\boldsymbol{S}^{-1}(\boldsymbol{x}^s) = \boldsymbol{B}^{-1} + \boldsymbol{H}(\boldsymbol{x}^s)^T \boldsymbol{R}^{-1} \boldsymbol{H}(\boldsymbol{x}^s) \tag{3.2}$$

where $\boldsymbol{H}(\boldsymbol{x}^s)$ is the gradient matrix evaluated for the solution vector $\boldsymbol{x}^s$. This expression (3.2) assumes that $\boldsymbol{B}$ and $\boldsymbol{R}$ are correct. The forward model H is weakly non-linear as is seen in equation 3.1, but local linearity is assumed for calculation of the solution error covariance $\boldsymbol{S}$.

3.5 The 1D-Var Cost Function

The n-dimensional state vector for a given occultation is defined as $\boldsymbol{x} = (\boldsymbol{T}; \boldsymbol{q}; p_{\text{surface}})$, where the $\frac{n-1}{2}$ dimensional vectors $\boldsymbol{T}$ and $\boldsymbol{q}$ hold the temperatures (in K) and specific humidities (in g/kg) on all $(\frac{n-1}{2})$ 1D-Var retrieval levels and p_{surface} is the surface pressure (in hPa) at the location of the occultation. The purpose of the 1D-Var processing is to find a maximum

[1] Currently ECMWF provides operational data at 137 levels, and reanalysis data; ERA Interim (60 levels) and ERA5 (137 levels). By September 1 2019 ERA-I was discontinued and replaced with ERA5. In future versions of the IFS model the ECMWF presumably will increase its number of levels. The ROM SAF will adapt to these changes. For a given background level structure 1D-Var requires a corresponding background error covariance prescription which has to be explicitly specified in the calling script. I.e. the background error covariance has to be recalculated offline when new level structures are applied.

[2] Suppose l is the number of model levels, then $n = 2l + 1$

[3] In the context of 1D-Var we reserve the "A" letter for references to the ECMWF 4D-Var analysis.

likelihood estimate of $\boldsymbol{x}$, given a profile of m observations $\boldsymbol{y}^o$ of the refractivity and some a priori knowledge $\boldsymbol{x}^b$, also referred to as a "background". Both the observations and the background have associated uncertainties described by an observational error covariance matrix $\boldsymbol{R}(m,m)$ and background error covariance matrix $\boldsymbol{B}(n,n)$. In the case of Gaussian error distributions, the statistically optimal solution that simultaneously fits both the observation and the background to within their respective errors, is obtained by minimizing the cost function

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b) + \frac{1}{2}(\boldsymbol{y}^o - H(\boldsymbol{x}))^T\boldsymbol{R}^{-1}(\boldsymbol{y}^o - H(\boldsymbol{x})) \tag{3.3}$$

with respect to the state $\boldsymbol{x}$. Here, $H(\boldsymbol{x})$ is the non-linear forward operator mapping the atmospheric state vector into measurement space. H converts the n dimensional state vector, defined on the 1D-Var levels and surface, into a corresponding profile of refractivity given on the m refractivity levels. It should be noted that the observation error covariance matrix $\boldsymbol{R}$ also includes the errors due to the forward modelling of the atmospheric state $\boldsymbol{x}$. It is implicitly assumed that all errors are non-systematic, i.e. the algorithm assumes biases in observation or background data are zero.

The 1D-Var levels are defined as a set of $\frac{n-1}{2}$ ECMWF model levels. Hence, the first term in the cost function $J(\boldsymbol{x})$ is evaluated on the 1D-Var levels, whereas the second term is evaluated on the observed levels which are expressed in terms of geopotential heights.

It may be shown that the minimization of the cost function J is effectively a least-square fitting procedure with $2J$ approximately having a χ^2 distribution with m degrees of freedom [RD.17]. This is briefly discussed in Section 4.1.

3.6 Minimising the Cost Function

Ideally, the 1D-Var solution does not depend on the choice of minimization algorithm. However, the choice of algorithm may have practical consequences, e.g. for the processing time, the memory required during processing, or the risk for a non-convergent behavior.

The ROPP 1D-Var code allows the choice between two minimisers: The Levenberg-Marquardt algorithm or the minROPP minimiser, developed for ROPP. The 1DV- package employs the Levenberg-Marquardt minimiser which was recommended in [RD.24].

The minimum of the costfunction is found through Newtonian iterative solution of

$$\boldsymbol{J}''(\boldsymbol{x}^{i+1} - \boldsymbol{x}^i) = -\boldsymbol{J}' \tag{3.4}$$

where $\boldsymbol{x}^i$ and $\boldsymbol{x}^{i+1}$ are the i-th and $(i+1)$-th approximation of $\boldsymbol{x}$. $\boldsymbol{J}'$ and $\boldsymbol{J}''$ are the first and second derivatives of the cost function with respect to $\boldsymbol{x}$. These derivatives are given by

$$\boldsymbol{J}'(\boldsymbol{x}^i) = \boldsymbol{B}^{-1}(\boldsymbol{x}^i - \boldsymbol{x}^b) - \boldsymbol{H}(\boldsymbol{x}^i)^T\boldsymbol{R}^{-1}(\boldsymbol{y}^o - H(\boldsymbol{x}^i)) \tag{3.5}$$

and

$$\boldsymbol{J}''(\boldsymbol{x}^i) = \boldsymbol{B}^{-1} + \boldsymbol{H}(\boldsymbol{x}^i)^T\boldsymbol{R}^{-1}\boldsymbol{H}(\boldsymbol{x}^i)) \tag{3.6}$$

In the Levenberg-Marquardt approach [RD.14] the diagonal values of $\boldsymbol{J}''$ are modified as $J''_{ii} = J''_{ii}(1+\lambda)$, where λ is a positive scalar value. If $\boldsymbol{x}^{i+1}$ reduces the cost function value, λ is reduced for the next iteration. Conversely, if it found that $\boldsymbol{x}^{i+1}$ increases the cost function

value, λ is increased and the increment is recalculated. This procedure is repeated until the cost function value stops falling. There is a maximum number of iterations, "`n_iter`", currently set to 24, and practically never reached because the algorithm converges before this point.

For specifics of the minROPP minimiser see Refs. [RD.19] and [RD.18].

3.7 Observational/Background Error Covariances

The uncertainties of atmospheric variables (refractivity, temperature, specific humidity and pressure) quantify the expected deviations from their true values expressed in standard deviations and biases. The derived uncertainties that come with each product are calculated from the uncertainties of the input variables (measurements and background uncertainties). This section describes the construction of observation and background error covariance matrices, which contain the input uncertainties to the 1D-Var algorithm.

3.7.1 Observational error covariance

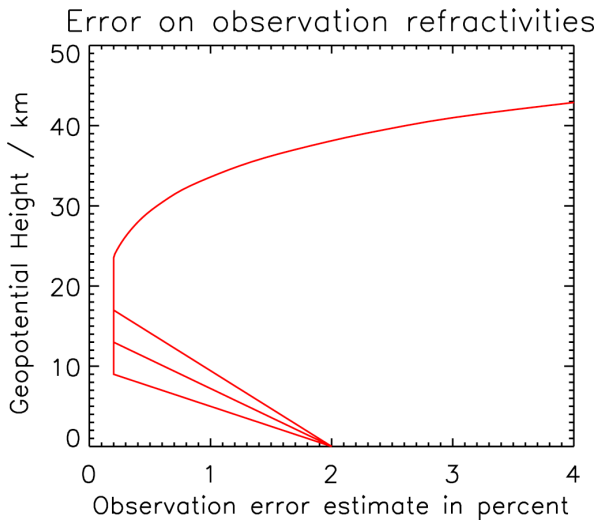


Figure 3.1: *Fractional refractivity error estimates for 3 different tropopause heights (9 km, 13 km, 17 km).*

The fractional refractivity errors vary with height, as at high levels the signal to noise ratio is decreasing with altitude. Similarly in the lower troposphere, particularly in the tropics, turbulence and horizontal gradients are introducing uncertainty. In the 1D-Var retrieval procedure, these issues are addressed through the error covariance matrix $\boldsymbol{R}$, which prescribes different weights to the data points through the observational errors.

Radio occultation errors in the troposphere are dominated by horizontal gradients, turbulence and the presence of water vapour (see for example [RD.11]). In the stratosphere, the circulation is stratified and almost zonally symmetric, so RO measurements are fairly accurate in this part of the atmosphere. Thus, the tropopause as boundary between stratosphere and troposphere is also the natural boundary between two different error regimes of radio occultation data. The error model employed in 1DV reflects this.

Figure 3.1 shows the observational error model that is implemented in 1DV. The refractivity

error estimate is 2% at the surface, decreasing linearly to 0.2% at the tropopause. Above the tropopause, the fractional errors are 0.2% or 0.02 N-units, whichever is greatest[4]. The error model is roughly resembling the refractivity error profiles reported in [RD.11] and [RD.29].

The altitude of the tropopause is determined from the background profile using the standard definition of the tropopause as the lowest level at which the lapse rate decreases to 2 K/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 K/km. This dynamically-determined tropopause height is then compared to a climatological value [RD.15]. This climatology is the mean tropopause height for the years 1958-1997 based on NCEP and radiosondes on a 2.5 × 2.5 degree grid. If the pressure at the dynamically-determined tropopause height differs by more than a factor of 2 from the climatological value the climatological value will be used instead of the dynamically-determined value.

The forward model uncertainty is included in the observation uncertainty. The $\boldsymbol{R}$ matrix is the sum of the observation error covariance and the forward model error covariance matrices. We cannot separate these two contributions to the $\boldsymbol{R}$ matrix. It is however possible to evaluate the validity of the $\boldsymbol{R}$ matrix assumption by looking at the Desrozier relations[RD.6], which predict certain statistical properties of a set of retrievals with common error covariance. As an example one Desrozier relation says that the $\boldsymbol{R}$ matrix is equal to $\left\langle(\boldsymbol{y}_o-\boldsymbol{y}_s)(\boldsymbol{y}_o-\boldsymbol{y}_b)^T\right\rangle$. This particular relation is illustrated in Figure 3.2. It is seen that diagonal of $\boldsymbol{R}$ is slightly underestimated around 20 km and overestimated around 10 km. We conclude that the $\boldsymbol{R}$ assumption is reasonable.

The observational error covariance matrix $\boldsymbol{R}$ used in 1DV is derived from the error estimates given above by the expression:

$$R_{ij} = \sigma_i \sigma_j e^{-|z_i - z_j|/h} \tag{3.7}$$

where σ_i and σ_j are the error estimates at two points i and j along a vertical profile, z_i and z_j are the geopotential heights and h is a characteristic scale for the decay of correlation. A uniform scale of h=3 km is used [RD.9], with the resulting correlation matrix shown in Figure 3.4 (right panel).

3.7.2 Background error covariance structure

The 1DV processing also needs the background error covariance matrix $\boldsymbol{B}$ as input. As was the case for the observation error covariances, this matrix is constructed from one-dimensional profiles of error estimates combined with an error correlation matrix $\boldsymbol{C}$:

$$B_{kl} = \sigma_k \sigma_l C_{kl} \tag{3.8}$$

The correlation matrix $\boldsymbol{C}$ is constant, and provided by ECMWF [RD.10]. The correlation matrix provided by ECMWF is given on 91 levels, while ERA-I background fields at 60 levels are used for CDR v1.0, ICDR v1.0, ICDR v1.1 and Offline v1.0. For the current NRT production ECMWF(OPER) at 137 levels is used, and for Offline v1.1 and onward, and for NTC, ERA5 at 137 levels is used. So we need to interpolate the correlation matrix to 60 and 137 levels. The procedure is as follows: The correlation matrix is interpolated as a 2 dimen-

[4] The index of refraction in the atmosphere is close to unity, so a given index of refraction, n, is conveniently expressed as $(n-1)10^6$ N-units

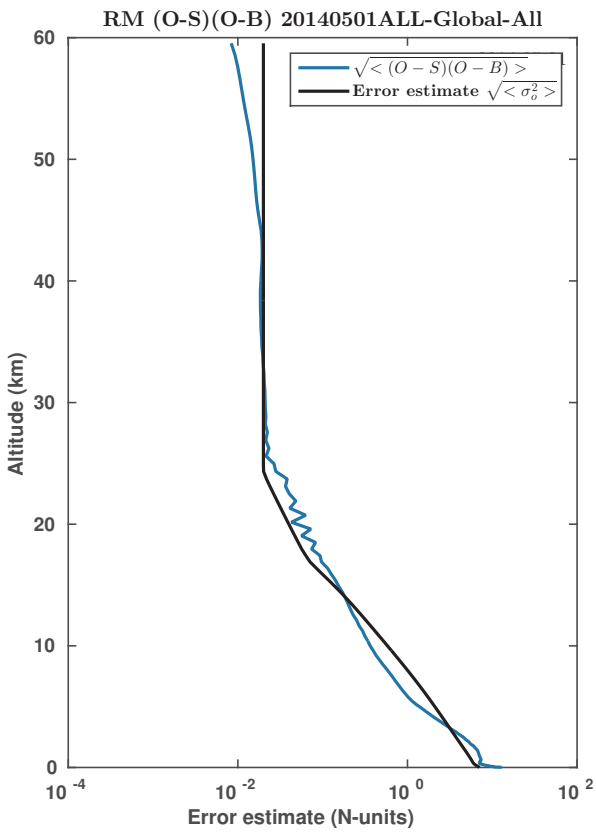


Figure 3.2: *Desrozier-plot, comparing the diagonal of* $\mathbf{R}$ *and* $<(\mathbf{y}_o - \mathbf{y}_s)(y_o - y_b)^T>$ *for a month of COSMIC profiles.*

sional function using splines. Temperature and specific humidity correlations are interpolated separately. The results are inspected visually. There is a discontinuity in the gradient at the diagonal, which leads to inaccurate interpolation at the diagonal. The resulting matrices are therefore forced equal to unity at the diagonals, and the temperature part, humidity part and the singular pressure diagonal point are combined to a full state vector correlation matrix. Finally the full correlation matrix is diagonalised and eigenvalues less than or equal to zero are forced to be equal to the lowest positive eigenvalue before reconstruction. This final step is performed to ensure positive definiteness. The error correlations are shown in Figure 3.3. It is assumed that there are no correlations between temperature and specific humidity and between surface pressure and specific humidity errors (see Figure 3.4 left panel).

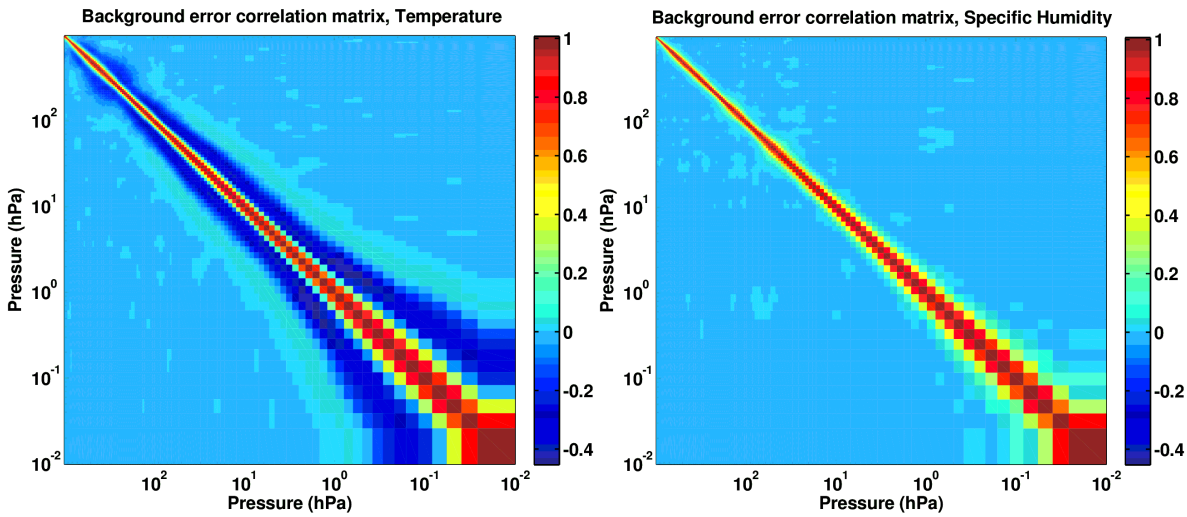


Figure 3.3: *Vertical correlations for the background errors given on the 137 model levels plotted on a set of pressures for a profile with the top level at 0.01 hPa and the bottom level at 1012 hPa: T error correlations in the left panel and q error correlations in the right panel. It is assumed that there are no correlations between temperature errors and humidity errors.*

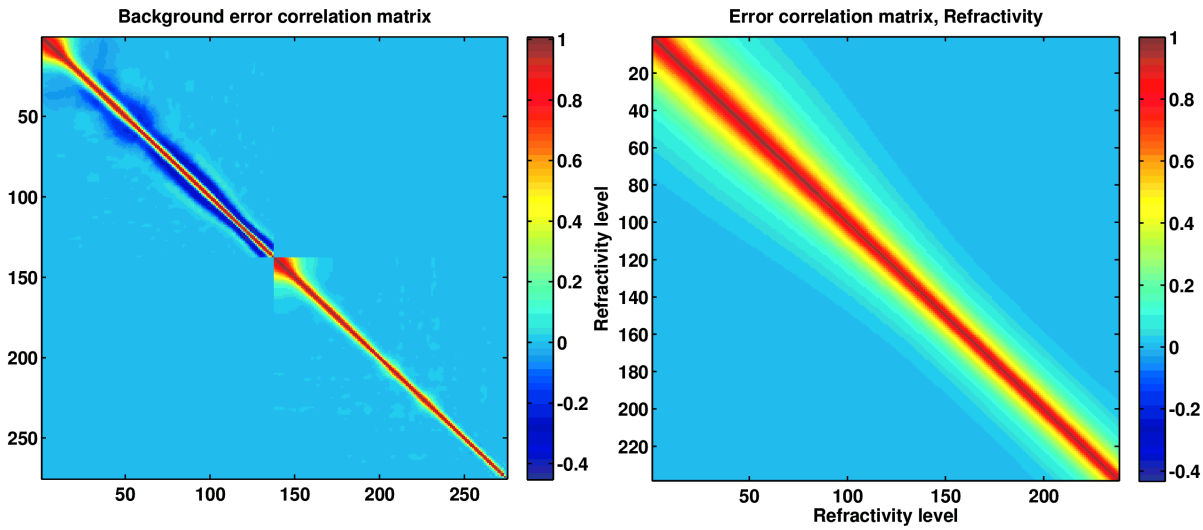


Figure 3.4: *Left: Error correlation matrix between 275 state-vector entries. Index 1-137; temperature error correlations. Index 138-274; specific humidity error correlations. Last row and column (index 275) shows surface pressure error correlations (zeros except for the diagonal). Right: Refractivity error correlation matrix.*

The assumed forecast error standard deviations depend on the product. These are described in chapter 6.1.

3.8 Reduction of Background Information Error

The "prior fraction" illustrates how 1D-Var reduces the a priori error through a retrieval. The prior fraction is the ratio between the theoretical error STDV (square root of equation 3.2) and the background error STDV for a retrieved property. In the upper left panel Figure 4.1

the prior fractions for the CDR v1.0 retrievals based on ECMWF ERA-I background fields are shown as an example. The plot basically illustrates where 1D-Var is adding information to the already known background profiles for a given configuration; When the prior fraction is close to zero the measurements are weighted high in the retrieval, and when the prior fraction is close to unity the background is predominantly weighted.

4 Practical Considerations

4.1 Evaluating the Retrieved Solution

The similarities with least-square fitting suggests that the cost function value at convergence, $J(\boldsymbol{x})$, can be used to validate the goodness of the fit. If the errors specified by the $\boldsymbol{R}$ and $\boldsymbol{B}$ matrices are in accordance with the actual errors then $2J$ is expected to follow a χ^2_m (chi-squared distribution with m degrees of freedom) where m is the dimension of observation space, i.e. m is the number of refractivity values used in the 1D-Var processing [RD.14]. Thus the expected value of $2J$ at convergence is m with a standard deviation of $\sqrt{2m}$. This is a subject of the validation report.

Additionally there are certain theoretical relations [RD.6] between the statistical properties of the refractivities and the assumed observation and background error covariances, that have to be valid if the applied error covariances are chosen properly. For example the matrix relation $\langle(\boldsymbol{y}^o - H(\boldsymbol{x}^b))(\boldsymbol{y}^o - H(\boldsymbol{x}^b))^T\rangle = \boldsymbol{O} + \boldsymbol{HBH}^T$ will be true if the data are unbiased and the $\boldsymbol{O}$ and $\boldsymbol{B}$ truly represent the actual observation and background errors. Such relations may help to verify the method in two ways. First, by constructing surrogate data sets with known error-covariances one can verify that the algorithm is actually performing correctly. We have done that and verified that the algorithm is sound in that sense. Secondly the matrix relations gives an indication about whether the error and (zero) bias assumptions are sound in the actual retrieval (see for instance in Figure 3.2).

4.2 Validating the Retrieved Solution

The 1D-Var products have be validated against ECMWF analysis. The criteria for validation are described in [AD.3]. Validation of a given product is an examination of plots showing the products ability to obey the threshold, target and objective limits described in [AD.3]. A few examples of such validation plots are shown for a month of COSMIC data processed with ERA-I background (Figure 4.1). These are just shown as examples. An actual discussion of the 1DV performance can be found in the validation reports for the various products. However, it should be noted that the reason for the apparent bias in pressure is discussed in section 6.2.4.

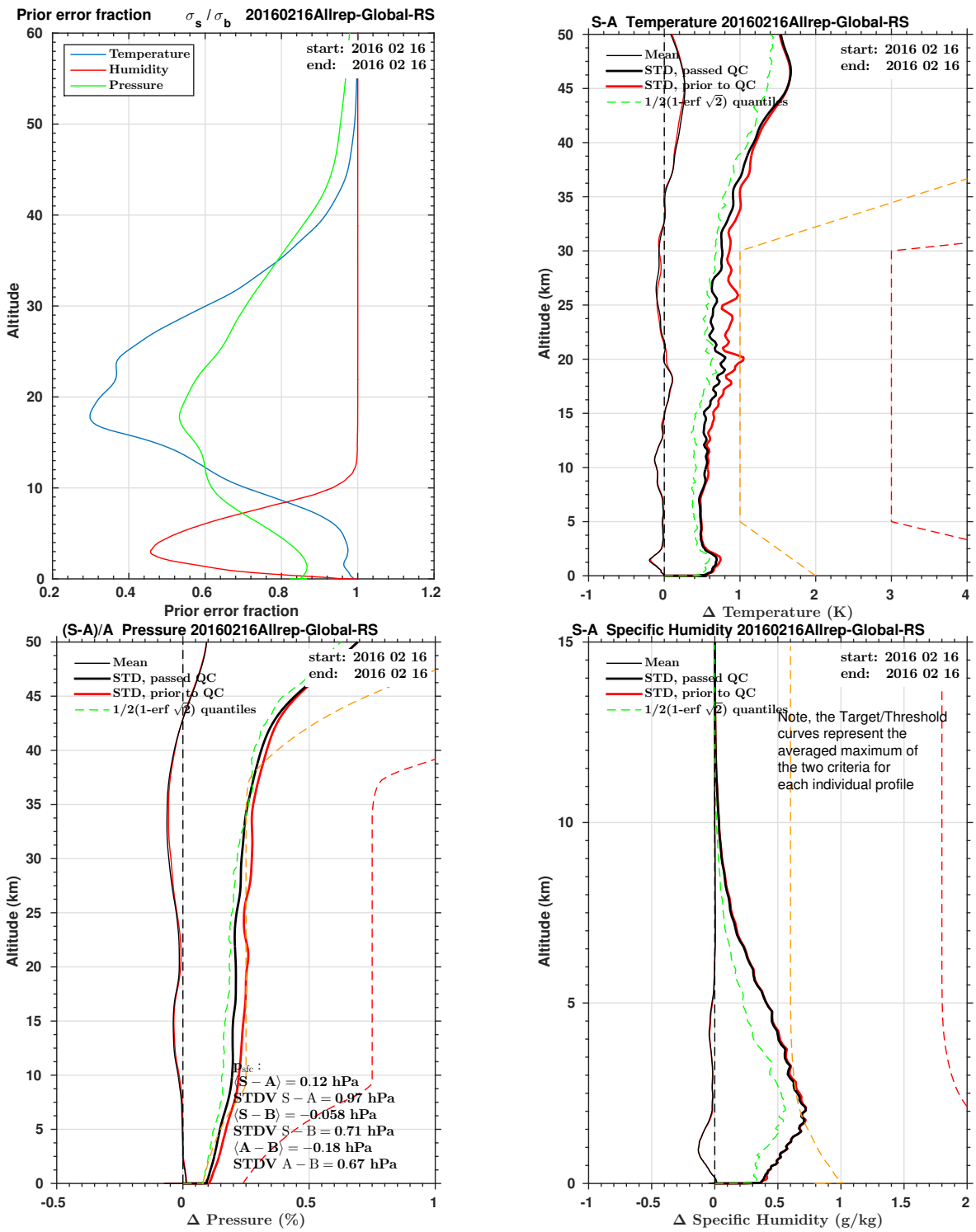
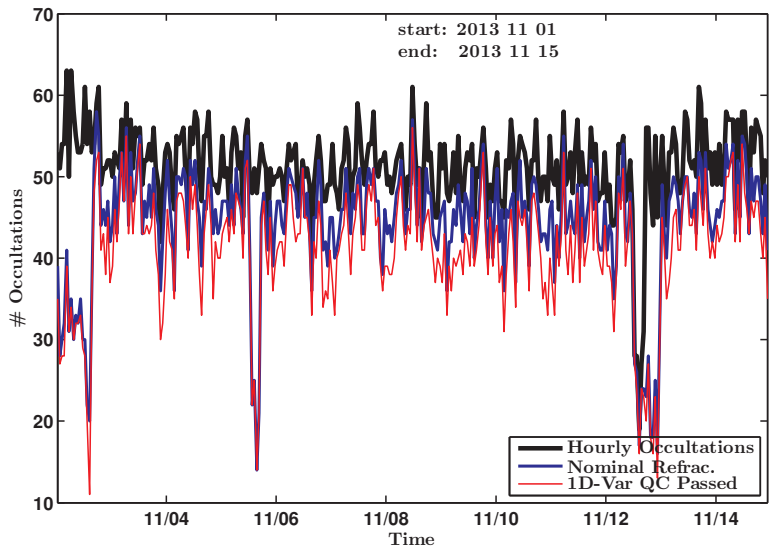


Figure 4.1: *Upper left: Example of prior fraction in the case of ERA-I background fields; Next: A few examples of validation-type plots, temperature (upper right), pressure (lower left) and specific humidity (lower right), here for Metop A/B data Dec 15 2016 and ERA-I background. The dashed red and orange curves show the averaged PRD thresholds and targets.*

4.3 Quality control

The input data for the 1D-Var processing, i.e. the refractivity profiles in the current configuration, comes with a quality control flag, the "PCD flag", which is a 16 bit integer containing quality QC results for different processing steps [RD.19]. The PCD-flag provided in the refractivity files contains a bit which is activated in case of non-nominal refractivity data. Solutions based on non-nominal refractivity profiles are produced but flagged as non-nominal in the further processing.

After the 1D-Var processing a further quality check is performed in order to flag out solutions on basis of the normalized cost function $2J/m$: If $2J/m \leq$ `J_s_limit` the profile is accepted[1]. Likewise the number of iterations used by the 1D-Var minimizer is used for a quality check. If the number of iterations is more than 25 then the retrieval is flagged out. These quality checks are merged into the non-nominal bit of the PCD flag which is then stored in the output file. An example of the performance of the QC on a 2 weeks period from November 2013 is illustrated in Figure 4.2. These NRT data are taken from the Metop A and B satellites (PPF 2.20).



QC performance; 20131101data-ses-Global-All

Total # of profiles:	15979
Profiles with nominal refractivity:	13810
Profiles that passed 1D-Var QC:	12849

Figure 4.2: *An example of QC performance. "Hourly occultations" means the number of Level 2A profiles per hour. Out of these a number of files with "nominal" refractivity are identified. The profiles that did not "pass 1D-Var QC" also include profiles where it was not possible to produce a refractivity profile with error estimates, meaning that 1D-Var was actually never executed in these cases.*

[1] In the current NRT product `J_s_limit` = 2, and in the CDR1, ICDR1, Offline and NTC processing `J_s_limit` = 5.

4.4 Exception handling

In all cases 1DV will attempt to produce an output. So the exceptions can be divided into two groups; those where an output "wet-file" is produced but corrupted, and those where the processing chain has been broken at some step, and output does not exist. Table 4.1 shows the exceptions that are handled explicitly in the 1D-Var production.

In the first case the output file, containing incomplete and/or non-nominal estimates of the temperature, specific humidity, pressure and geopotential height will contain an activated PCD flag indicating "Meteorological processing non-nominal".

In the second case 1DV will return an exit code, containing information about where and why the processing was terminated. These exit codes are interpreted by the operational (NRT) monitoring system.

ERROR 12	Background file does not exist.	No output.
ERROR 30	Missing thinned refractivity file.	No output.
ERROR 31	Too few valid refractivity data points.	No output.
ERROR 51	Observation and background data files must contain an equal number of profiles.	No output.
ERROR 59	Observation covariance is not OK (could be a missing cor-file).	No output
ERROR 52	Too many iterations.	Output exists.
ERROR 53	$\frac{2J}{m}$ greater than, `J_s_limit`.	Output exists.

Table 4.1: *Exceptions that are registered in the operational (NRT) monitoring system.*

4.5 Outputs

The 1D-Var algorithm outputs a "wfm"-file for each occultation containing the atmospheric variables and their estimated uncertainties and occultation information / meta data inherited from the observation file. The "wfm"-file also contains forward propagated bending angle and refractivity with error-estimates, but these parameters are removed in the disseminated "wet"-file. In table 4.2, all variables in a "wfm"-file are listed.

Table 4.2: *Full list of variables in an "wfm" file, as produced by ROPP 8.1 . Note that the file also contains empty spaces for derived variables, like tropopause parameters, which are not populated at this stage.*

Shortname	**Longname**
occ_id	Occultation ID
gns_id	GNSS satellite ID
leo_id	LEO satellite ID
stn_id	Ground station ID
start_time	Starting time for the occultation
year	Year
month	Month

Continued on next page ...

Table 4.2: *Full list of variables in an "wfm" file, as produced by ROPP 8.1 . Note that the file also contains empty spaces for derived variables, like tropopause parameters, which are not populated at this stage.*

Shortname	**Longname**
day	Day
hour	Hour
minute	Minute
second	Second
msec	Millisecond
pcd	Product Confidence Data
overall_qual	Overall quality
time	Reference time for the occultation
time_offset	Time offset for georeferencing (since start of occ.)
lat	Reference latitude for the occultation
lon	Reference longitude for the occultation
undulation	Geoid undulation for the reference coordinate
roc	Radius of curvature for the reference coordinate
r_coc	Centre of curvature for the reference coordinate
azimuth	GNSS->LEO line of sight angle (from True North) for the reference coordinate
alt_refrac	Geometric height above geoid for refractivity
geop_refrac	Geopotential height above geoid for refractivity
refrac	Refractivity
refrac_sigma	Estimated error (1-sigma) for refractivity
refrac_qual	Quality value for refractivity
dry_temp	Dry temperature
dry_temp_sigma	Estimated error (1-sigma) for dry temperature
dry_temp_qual	Quality value for dry temperature
geop	Geopotential height above geoid for P,T,H
geop_sigma	Estimated error (1-sigma) for geopotential height
press	Pressure
press_sigma	Estimated error (1-sigma) for pressure
temp	Temperature
temp_sigma	Estimated error (1-sigma) for temperature
shum	Specific humidity
shum_sigma	Estimated error (1-sigma) in specific humidity
meteo_qual	Quality value for meteorological data
geop_sfc	Surface geopotential height
press_sfc	Surface pressure
press_sfc_sigma	Estimated error (1-sigma) for surface pressure
press_sfc_qual	Surface pressure quality value
Ne_max	Peak Chapman layer electron density
Ne_max_sigma	Est error in peak Chapman layer electron density
H_peak	Height of Chapman layer peak
H_peak_sigma	Est error in height of Chapman layer peak
H_width	Width of Chapman layer

Continued on next page ...

Table 4.2: *Full list of variables in an "wfm" file, as produced by ROPP 8.1 . Note that the file also contains empty spaces for derived variables, like tropopause parameters, which are not populated at this stage.*

Shortname	**Longname**
H_width_sigma	Est error in width of Chapman layer
tph_bangle	Bending angle-based TPH
tpa_bangle	Bending angle-based TPA
tph_bangle_flag	Bending angle-based TPH QC flag
tph_refrac	Refractivity-based TPH
tpn_refrac	Refractivity-based TPN
tph_refrac_flag	Refractivity-based TPH QC flag
tph_tdry_lrt	Dry temperature-based TPH (lapse rate)
tpt_tdry_lrt	Dry temperature-based TPT (lapse rate)
tph_tdry_lrt_flag	Dry temperature-based TPH QC flag (lapse rate)
tph_tdry_cpt	Dry temperature-based TPH (cold point)
tpt_tdry_cpt	Dry temperature-based TPT (cold point)
tph_tdry_cpt_flag	Dry temperature-based TPH QC flag (cold point)
prh_tdry_cpt	Dry temperature-based PRH (cold point)
prt_tdry_cpt	Dry temperature-based PRT (cold point)
prh_tdry_cpt_flag	Dry temperature-based PRH QC flag (cold point)
tph_temp_lrt	Temperature-based TPH (lapse rate)
tpt_temp_lrt	Temperature-based TPT (lapse rate)
tph_temp_lrt_flag	Temperature-based TPH QC flag (lapse rate)
tph_temp_cpt	Temperature-based TPH (cold point)
tpt_temp_cpt	Temperature-based TPT (cold point)
tph_temp_cpt_flag	Temperature-based TPH QC flag (cold point)
prh_temp_cpt	Temperature-based PRH (cold point)
prt_temp_cpt	Temperature-based PRT (cold point)
prh_temp_cpt_flag	Temperature-based PRH QC flag (cold point)
level_type	Vertical level type
level_coeff_a	Hybrid / Eta level coefficient (a or eta)
level_coeff_b	Hybrid / Eta level coefficient (b or tau)
J	Cost function value at convergence
J_scaled	Scaled cost function value at convergence (́2J/ḿ)
J_init	Initial Cost function value
n_iter	Number of iterations
minropp_exit_mode	Exit mode of minimiser (minROPP)
pge_gamma	Probability of Gross Error "gamma" value
J_bgr	Background contribution to cost function profile
J_obs	Observation contribution to cost function profile
OmB	Observation - Background refractivity
OmB_sigma	Expected (Observation - Background refractivity) standard deviation
OmA	Observation - Analysis refractivity
OmA_sigma	Expected (Observation - Analysis refractivity) standard deviation

Continued on next page ...

Table 4.2: *Full list of variables in an "wfm" file, as produced by ROPP 8.1 . Note that the file also contains empty spaces for derived variables, like tropopause parameters, which are not populated at this stage.*

Shortname	Longname
B_sigma	Background refractivity standard deviation
pge	Probability of Gross Error
AmB	H(x)-H(xb) in refractivity space

4.5.1 Example of retrieved meteorological fields

For illustration we show an example of a retrieved profile in this section. The data used are from RE1B, Metop B, December 2016. In Figure 4.3, the red curves show the background temperature and specific humidity profiles, and the blue curves show the solution profiles. The temperature is hardly modified by the refractivity data in the troposphere while the humidity is actually changed quite a bit in this case.

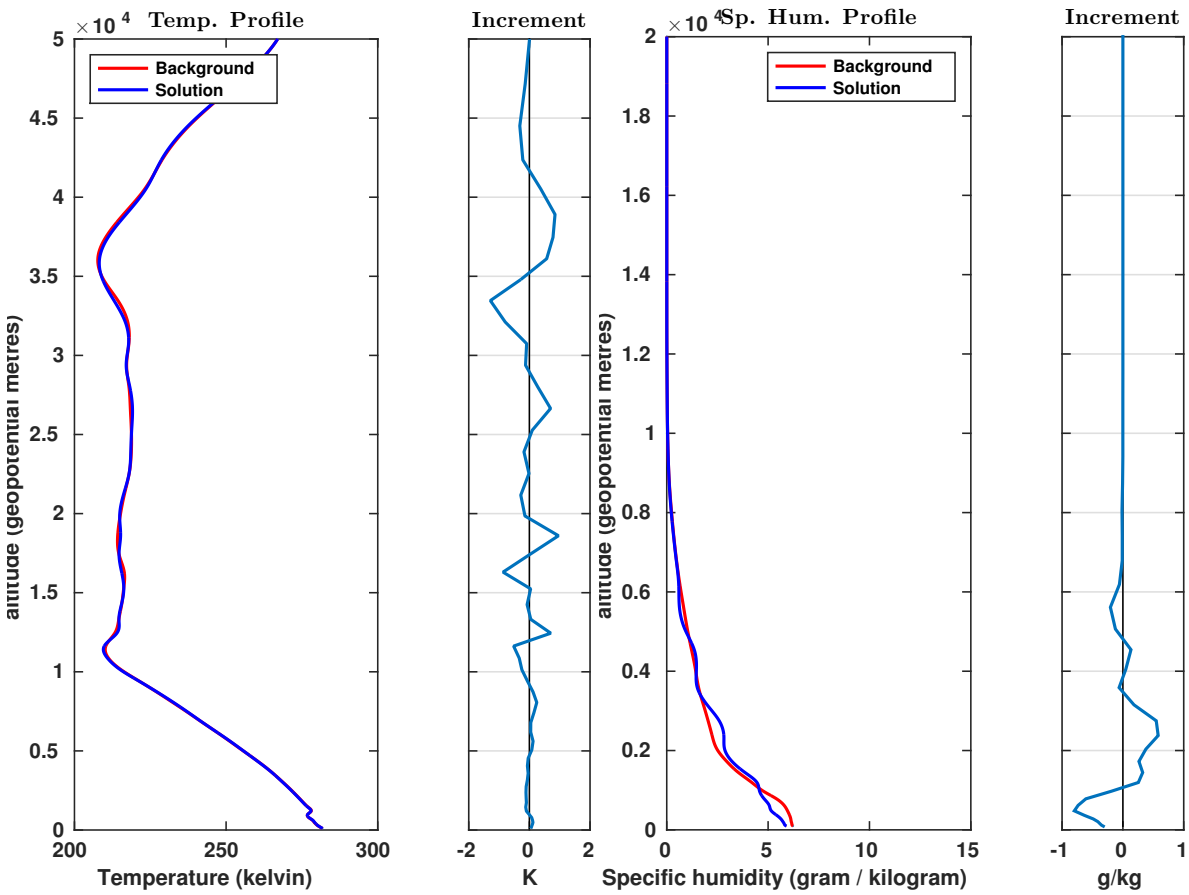


Figure 4.3: *Temperature (left) and specific humidity (right) profiles from a single occultation on December 15, 2016, UTC 12:35, Latitude -55.55 deg. and Longitude -1.69 deg. The red curves show the background assumption used and the blue curves show the solution which has incorporated the information from the retrieved refractivity profile. The right columns show the increments of the meteorological parameters for this particular profile.*

The temperature increment is largest in the stratosphere while the specific humidity incre-

ment is largest in the troposphere. This is due to the background covariance structure, written in the $\boldsymbol{B}$-matrix. In the stratosphere the background specific humidity variances and covariances are small and the humidity contribution to refractivity is below the observation noise, which means that the retrieved specific humidity is forced to follow the background specific humidity, which happens to be practically equal to zero. In the troposphere on the other hand the background specific humidity uncertainty at a given level is large with respect to the "equivalent" temperature background uncertainty at the same level (when both uncertainties are mapped into refractivity space and compared there).

5 Assumptions and Limitations

5.1 Assumptions

The most critical assumptions behind the 1D-Var processing are assumptions of the covariance structures in observation space as well as state space. The covariances may not be representing the true errors properly, and that may introduce errors in the solutions. Sub grid stochastic errors are not represented, and the influence of assuming unbiased background fields on the 1D-Var results is not taken into account.

The overall balance between measurement information and background information of the 1D-Var products is determined by the assumed measurement noise and the assumed background noise. The two matrices have been constructed in very different ways with different techniques. Their mutual scaling could possibly be optimized, and this issue still has to be investigated further by consistency checks of the statistical properties of the retrieval products.

5.2 Limitations

The 1D-Var products have the highest information content in the Upper Troposphere / Lower Stratosphere while the retrieval errors grow in the lower troposphere and upper stratosphere. In the upper stratosphere the refractivity is low compared to the noise level and that causes 1D-Var to fall back on the background above approximately 40 km.

Not all data points in a profile are equally representative of the true refractivity. The fractional errors vary with height as illustrated in Figure 3.1, and at high levels biases, e.g. those due to the model-dependent initialization of the Abel transform, may become significant. Similarly, at the lowest levels, particularly in the tropics, a negative refractivity bias related to super refraction and turbulence is causing a negative bias in the specific humidity.

As seen in section 4.5.1 the background fields are woven into the meteorological products in subtle ways. There is a gradual transition from levels where the measurements have an impact to levels where the products basically repeat the ECMWF-fields. It is of importance to communicate the actual information content of the 1D-Var products to the users.

Generally the specific humidity does not contain any observational information above the tropopause[1], meaning that the reported specific humidity in the stratosphere is basically just the background model specific humidity. See Figure 4.1 upper left panel.

Likewise the temperature does not contain any observational information in the mid and lower troposphere, meaning that the reported temperature below approximately 7 km is basically just the background model temperature. See Figure4.1 upper left panel.

The errors introduced by the assumptions of the forward model, especially the spherical

[1] In CDR1, ICDR1 and Offline 1.0 a bug in the forward model forces the specific humidity to be larger that 10-6 g/kg in the evaluation of the humidity. This cannot be fixed by definition, because the ICDR1 has to proceed with the same algorithm as CDR-v1.0. In Offline-v1.1 the bug should have been fixed but we did not mange to complete tests before the code was frozen. Consequently the 1D-Var output in Offline-v1.1 is allowed to duck below zero, but the optimization is based on a forward model which puts q = 1e-6 g/kg. In NTC, starting with Sentinel-6 data, the forward model has been fixed such that q is mapped linearly into refractivity.

symmetry assumption, are not quantified. In areas where horizontal gradients are present, this assumption is violated, and will be a source of errors in the 1D-Var products.

The Offline, NTC and NRT products are subject to upgrades from time to time, and this will cause time series from these products to be unstable in terms of accuracy, and bias jumps are expected. These products are not suitable for trend studies.

Finally it should be noted that even though a single receiver covers the whole globe over time, the diurnal cycle is not sampled homogeneously. This limitation is described and dealt with in the Level 3 data production [RD.22]. A user who wants to use profile data for climate studies should be aware of this issue and take it into account.

6 Description of differences for NRT, Offline, NTC, CDR1 and ICDR1

This section describes the specific configuration choices for the NRT, Offline, NTC, CDR and ICDR products. The current reprocessing configuration has been developed regarding the definition of background profiles and background profile uncertainty, thus it has departed substantially from the NRT configuration. The current Offline configuration is based on the ICDR configuration, but further developments has been and will be implemented in the Offline processing. In future NRT upgrades the NRT configuration will be brought closer to the offline configurations, and the present document will be updated accordingly. The NTC production is configured in the same way as the Offline production except for the background fields, which are ECMWF operational forecasts interpolated in space and time. The NTC code contains an upgraded version of the forward moel, where specific humidity is mapped linearly into refractivity.

The following section (6.1) describes the NRT configuration, while the next sections (6.2, 6.3 and 6.4) explains where the data products CDR-v1.0, ICDR-v1.0, Offline-v1.0, ICDR-v1.1, Offline-v1.1 and NTC configurations deviate from the NRT configuration.

6.1 Configuration for ROM SAF 1D-Var, NRT products

6.1.1 Observations from Metop 1st and 2nd generation

RO profiles, distributed across the globe, are continuously generated from the receivers sitting on the Metop satellites. The profiles are processed to bending angles by EUMETSAT, and further to Level 2A (refractivity) in the ROM SAF.

These refractivity profiles are given on a set of geopotential heights, from near the surface to around 80 kilometers height. The 1D-Var NRT products are based on refractivity profiles. Before being passed to the 1D-Var processing the observation profiles are thinned to a set of 247 heights based on a set of 247 underlying fixed impact heights. For reference to the selection of impact height levels see [RD.23].

6.1.2 Background profiles from ECMWF Forecasts

For each occultation the background state is taken from an ECMWF global forecast from the standard dissemination stream. This forecast has been specified on 137 model levels since ECMWF introduced IFS Cycle Cy38r2 on 25 June 2013. The model levels are defined by the pressures at a set of intermediate half levels, and the pressures at the full levels (i.e. the model levels) are then obtained as the mean of the pressures at the two nearest half levels [1]. The other state variables ($\boldsymbol{T}$, $\boldsymbol{q}$, etc.) are given on the full model levels. The pressures at the intermediate half levels are defined by

$$p_{k+\frac{1}{2}} = a_{k+\frac{1}{2}} + b_{k+\frac{1}{2}} p_{\mathrm{sfc}} \tag{6.1}$$

where p_{sfc} is the model surface pressure and $a_{k+\frac{1}{2}}$ and $b_{k+\frac{1}{2}}$ are a set of constant coefficients [RD.7].

[1] In the CDR, ICDR and offline productions an alternative method is used to calculate full level pressures, see section 6.2.2

Each occultation has latitude, longitude, and time associated with it. For the background state, one selects the newest ECMWF forecast at the closest 6-hourly time step (i.e. at UTC 0, 6, 12, 18) taking care to avoid forecasts which will already have assimilated the observed RO profile. The ECMWF forecasts ("oper" stream) starts from the analysis at 0 UTC and 12 UTC. The 6 hour 4D-Var analysis windows ranges from 3 UTC to 9 UTC and from 21 UTC to 3 UTC. Hence the forecasts used in the nominal situation are as shown in table 6.1.

No time interpolation is performed in NRT production. The nearest neighbour field is chosen in time domain. The model fields are linearly spatial interpolated to the same geographic location as the observed profile. For each observed profile we thus obtain a corresponding model profile which provides the a priori information that is required for the 1D-Var analysis.

0:00 UTC - 3:00 UTC:	forecast valid at 0 UTC (step 12 of forecast started at	12 UTC*)
3:00 UTC - 9:00 UTC:	forecast valid at 6 UTC (step 6 of forecast started at	0 UTC)
9:00 UTC - 15:00 UTC:	forecast valid at 12 UTC (step 12 of forecast started at	0 UTC)
15:00 UTC - 21:00 UTC:	forecast valid at 18 UTC (step 6 of forecast started at	12 UTC)
21:00 UTC - 24:00 UTC:	forecast valid at 0 UTC next day (s. 12 of f. st. at	12 UTC)

Table 6.1: *The table shows how forecast times are attributed to occultation times. The term "step" refers to number of hours after the analysis time where the forecast where initiated. (*) Here the forecast is initiated on the day before.*

6.1.3 Mean background uncertainty estimates from Scaled Ensemble Standard Deviation

The forecast error standard deviation estimates for temperature and specific humidity, i.e. the σ_ν's in equation (3.8), are based on the "Scaled Ensemble Standard Deviation" (acronym: "ses") variable[2], which quantifies the error as the difference between ECMWF forecast and analysis. Here we specifically use time averages of the "ses" products from the year 2012, i.e. the error model is constant in time. The "ses" uncertainty estimates were disseminated at 91 levels in 2012, and here we apply a direct interpolation of the standard deviation profiles to 137 levels. The temperature uncertainty is assumed to be a constant profile, valid at all latitudes and longitudes, found by averaging over uncertainty fields from 36 days spread over the year 2012. In Figure 6.1 the temperature uncertainty is plotted as function of time and latitude, and in Figure 6.2 the specific humidity uncertainty is plotted as function of time and latitude. The specific humidity has large latitudinal variability. By assuming a constant relative uncertainty profile the variability in specific humidity uncertainty is to some extent represented, because the specific humidity itself shows a similar temporal and spatial pattern (see Figure 6.3. I.e the specific humidity uncertainty at level k, σ_k^q, calculated individually for each profile, is found as

$$\sigma_k^q = \frac{q_k^b}{\overline{q_k}} \overline{\sigma_k^{q,\mathbf{ses}}} \tag{6.2}$$

[2] Specifically, the data type "ses" (stream=enda, class=od) disseminated by ECMWF, representing the estimated uncertainty of the forecast at 0900 UT and 2100 UT is retrieved. The scaled ensemble STD represents the spread between temperature forecasts among the members of the ensemble data assimilation system, re-scaled to match the typical difference between forecasts and analyses. If anything, the "ses" field underestimates the background uncertainty at step 6, 12 and 18.

where q_k^b is the background specific humidity for the profile, $\sigma_k^{q,\mathbf{ses}}$ is the specific humidity error of first guess published by ECMWF and q_k is corresponding forecast of specific humidity. The bars mean average taken over the uncertainty fields and forecast fields corresponding to the same (36×2) times as were used for the temperature uncertainty estimate. In Figure 6.4 the quality of this approximation is shown in a scatter plot comparison of the disseminated specific humidity error versus the uncertainty estimate obtained by scaling the relative uncertainty with the background humidity profile for a single global snapshot. The relative uncertainty of the surface pressure is calculated directly as the mean of the error standard deviation of the logarithm to the surface pressure

$$\overline{\sigma^{p_{\mathrm{srf}}}/p_{\mathrm{srf}}} = \overline{\sigma^{\ln p_{\mathrm{srf}},\mathbf{ses}}} = 0.772 \times 10^{-4} \tag{6.3}$$

That is, a surface pressure of e.g. 1000 hPa would correspond to an uncertainty of 0.772 hPa.

The reason for these choices is that the temperature error of first guess is rather homogeneous, while humidity errors vary several orders of magnitude both as function of altitude and latitude. By assuming the relative uncertainty of the specific humidity to be constant we allow for a large variability of the specific humidity uncertainty.

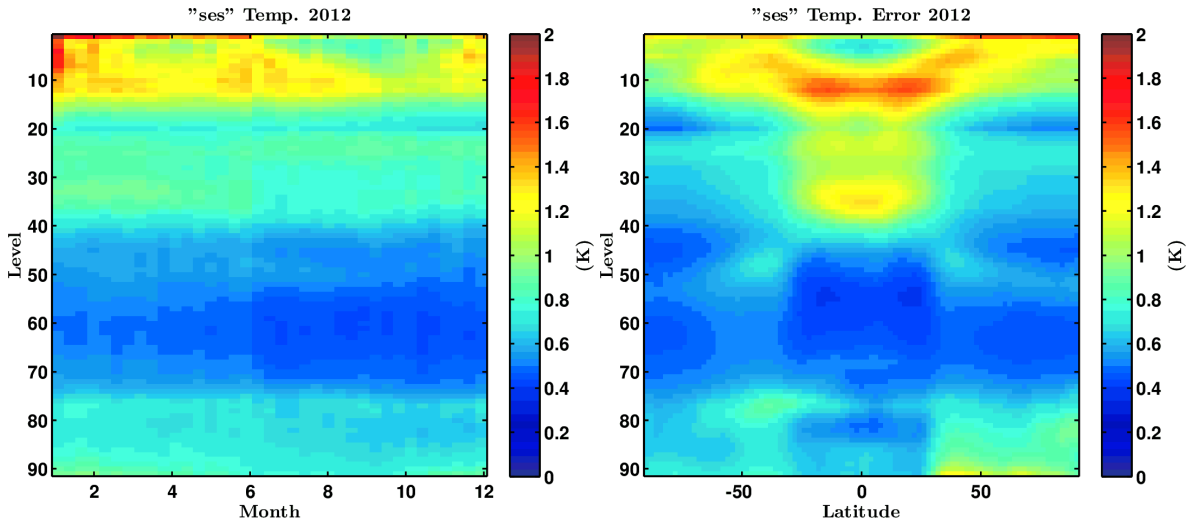


Figure 6.1: *Error estimate of temperature ('***ses***) as function of time (left) and latitude (right) during the year 2012.*

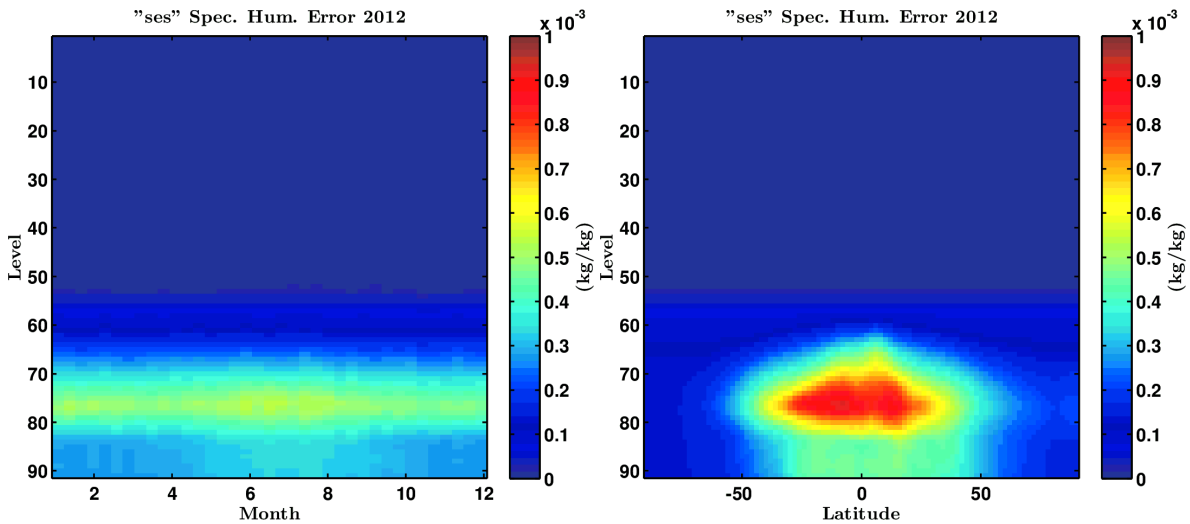


Figure 6.2: *Error estimate ('***ses***) of ECMWF specific humidity forecast as function of time as (left) and zonal mean over the year 2012 (right).*

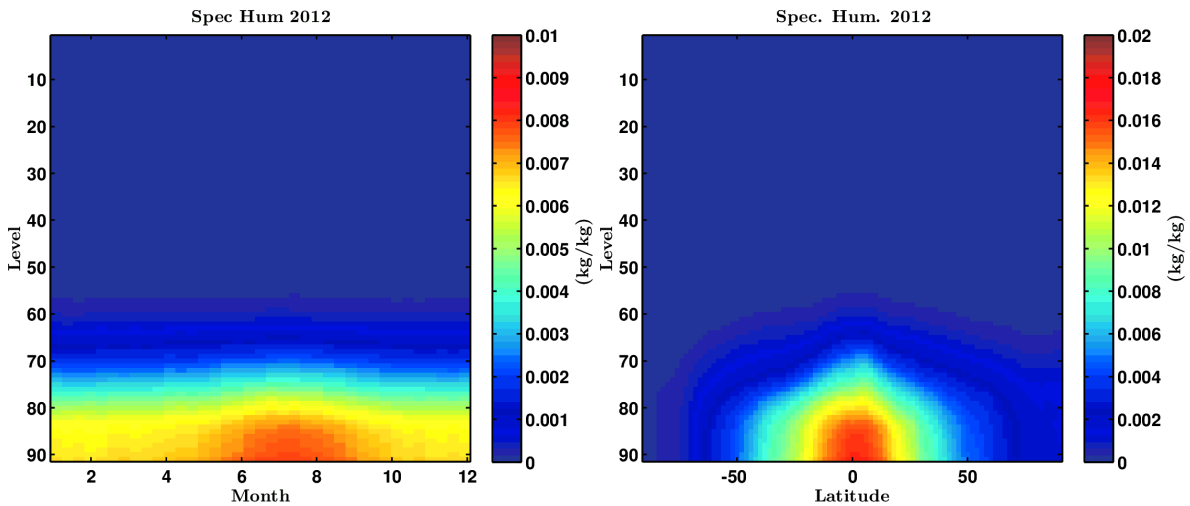


Figure 6.3: *Specific humidity as function of time (left) and latitude (right) during the year 2012.*

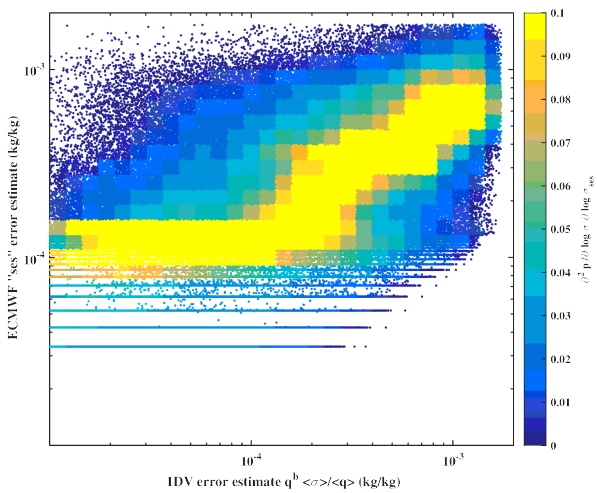


Figure 6.4: *Scatter plot of the ROM SAF specific humidity assumed background error versus the original ECMWF "ses" error estimate. The coloring shows the probability density in logarithmic space.*

The resulting temperature uncertainty and the uncertainty of specific humidity as function of level number are plotted in Figure 6.5. The Figure also shows two experimental uncertainty models which have been used historically, "ef" (error of first guess) and an old uncertainty estimate originally received from M. Fisher (2004), which are kept here for internal reference. Updates of the NRT background error covariance will occasionally be considered in response to future ECMWF upgrades. The code and configuration for the NRT production is collected in GPAC v 0042.

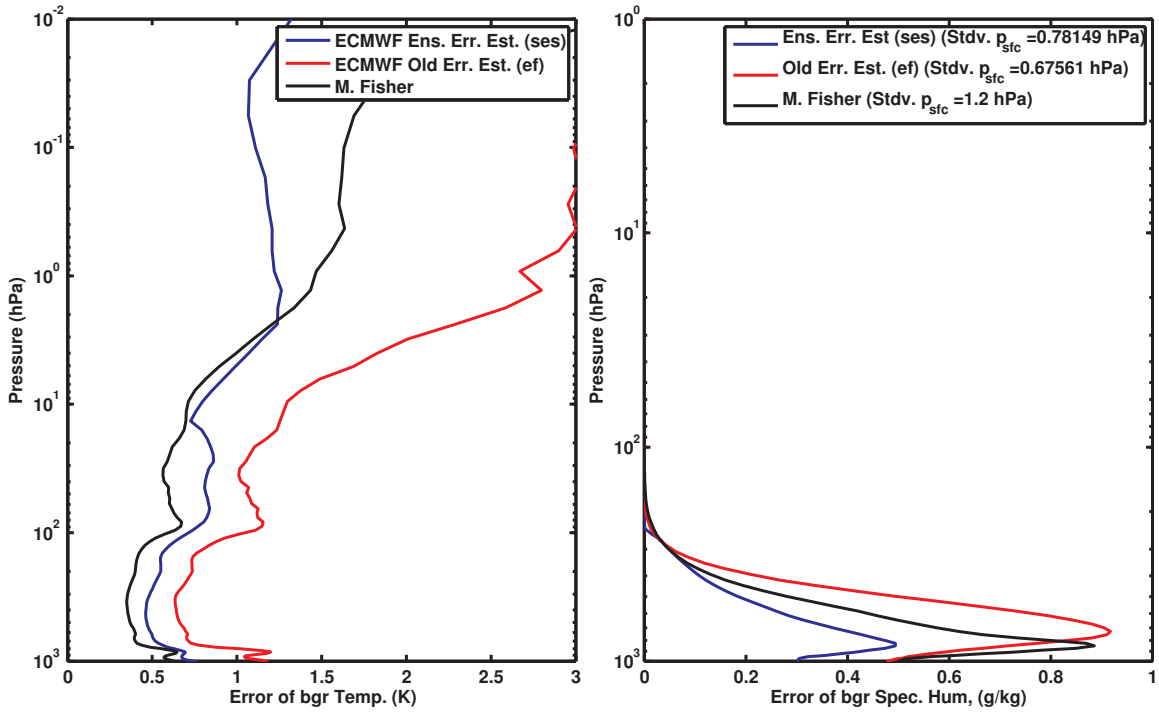


Figure 6.5: *Temperature background uncertainty profile (left) and specific humidity background uncertainty profile (right). (See legend for surface pressure background error). Specific humidity uncertainty is estimated by multiplying the relative uncertainty with the ECMWF global annual mean specific humidity. The blue curves represent the resulting uncertainty estimates globally averaged over 36 representative dates through out the year 2012. The Figure also shows the previous background uncertainty assumptions based on the so-called "ef" (error of first guess) fields and the old uncertainty estimate originally received from M. Fisher (2004).*

6.2 Configuration for first ROM SAF reprocessing, RE1 (1DV-3.3 and 1DV-3.4)

The first ROM SAF reprocessing, RE1, includes the climate data record CDR-v1.0 and interim climate data records ICDR-v1.0 and ICDR-v1.1.

6.2.1 Observations from multiple satellite missions.

The observation input data to CDR-v1.0 includes all available RO profiles from the Metop, COSMIC, CHAMP and GRACE missions. The ICDR-v1.0 and ICDR-v1.1 contains only Metop data. The Level 2A (refractivity) profiles are produced by the ROM SAF as described in [RD.27].

6.2.2 Background from ECMWF ERA Interim and ERA5

The background data until end of July 2019 has been retrieved from the ERA Interim forecast which is available with 3 hourly intervals up to 12 hours starting from the analysis at 0 UT and 12 UT. From August 1st 2019 the background profiles are retrieved from ERA5.

Transition from ERA-I to ERA5

By definition the ICDR-v1.0 and ICDR-v1.1 has to be based on the same code as the CDR-v1.0. Since ERA5 has 137 levels and ERA-I has only 60 levels a workaround has been implemented which is constrained to a preparation of the background data, such that the processing itself is unchanged. In CDR-v1.1 the 60 level background profile is prepared in the following way:

1. First half level and full level pressure at 138 and 137 levels are calculated from the ERA5 surface pressure.
2. Then then ERA5 surface pressure is used in combination with the (61) fixed half level coefficients of ERA-Interim to calculate a pseudo half level pressure array corresponding to the ERA-Interim level structure.
3. Next 60 corresponding full level pressures are calculated as described in Appendix IV (page 66).
4. Then interpolation of the 137 level temperature and humidity to 60 levels are performed, using a logarithmic pressure axis and cubic splines.
5. Finally the geo-potential height is recalculated at 60 levels using the new pressure, temperature and humidity vectors.

The prepared background profile is then ingested by 1D-Var exactly as an ERA-Interim profile.

Time interpolation of background profiles in CDR-v1.0, ICDR-v1.0, ICDR-v1.1 and Offline

Unlike the NRT background fields the profiles are produced by linear interpolation between the two nearest forecast steps. Table 6.2 shows which forecast steps are used in the interpolations for each time interval throughout the diurnal cycle.

Pressure half to full level interpolation in ERA-I

In contrast to the NRT case the full level pressures are calculated as suggested in [RD.30] eq. 3.18 for reprocessing. The reason for this change is that ERA-I appears to impose biases in the 1D-Var solution which are caused by incorrect attribution of pressure values at full levels. By applying an interpolation method which is consistent with the geopotential height calculation several solution bias issues in stratosphere and lower troposphere are resolved. For reference we summarize the impact of going from simple linear average of closest half levels to the method of [RD.30] in Appendix IV (page 66). The S&B Simmons and Burridge method was implemented during preparation of CDR-v1.0, and used in all following product updates.

00:00 UTC - 03:00 UTC:	12h forecast for 00 UTC comb. w.	3h forecast for 03 UTC
03:00 UTC - 06:00 UTC:	3h forecast for 03 UTC comb. w.	6h forecast for 06 UTC
06:00 UTC - 09:00 UTC:	6h forecast for 06 UTC comb. w.	9h forecast for 09 UTC
09:00 UTC - 12:00 UTC:	9h forecast for 09 UTC comb. w.	12h forecast for 12 UTC
12:00 UTC - 15:00 UTC:	12h forecast for 12 UTC comb. w.	3h forecast for 15 UTC
15:00 UTC - 18:00 UTC:	3h forecast for 15 UTC comb. w.	6h forecast for 18 UTC
18:00 UTC - 21:00 UTC:	6h forecast for 18 UTC comb. w.	9h forecast for 21 UTC
21:00 UTC - 24:00 UTC:	9h forecast for 21 UTC comb. w.	12h forecast for 24 UTC

Table 6.2: *List of forecasts attributed to time intervals in a diurnal cycle. The deterministic forecasts starting from the analysis at 0 and 12 UTC depends on data assimilated in a window starting 9 hours before and ending 3 hours after the analysis time. The ERA-I assimilation system includes RO data, and that implies that the background profiles in the intervals 0:00 UTC - 3:00 UTC and 12:00 UTC - 15:00 UTC are not independent of the observed refractivity in these intervals, because the 3 hour forecast contains information from RO profiles in these time intervals. The implications of this "contamination" problem is discussed in section 6.2.4*

6.2.3 Mean background uncertainty estimates from ERA-I error of first guess estimates

The ERA Interim data set comes with uncertainty estimates ("ef"-fields) valid at 3 hours forecasts. These are used to form a static temperature uncertainty and a static relative specific humidity uncertainty. The standard deviations are calculated from a set of ERA-I files ranging over the whole CDR-v1.0 period, including two GRIB files per month, one at UT03 and one at UT15 on the 15th of each month. These are 60 level data, but the error correlation matrix is the same as for NRT, interpolated to 60 levels. The reanalysis is done with another (older) IFS version, on only 60 levels, and larger values are found for the resulting temperature and specific humidity background error covariances for RE1. This is not the case for surface pressure though. The background error standard deviations are shown in Figures 6.6 and 6.7.

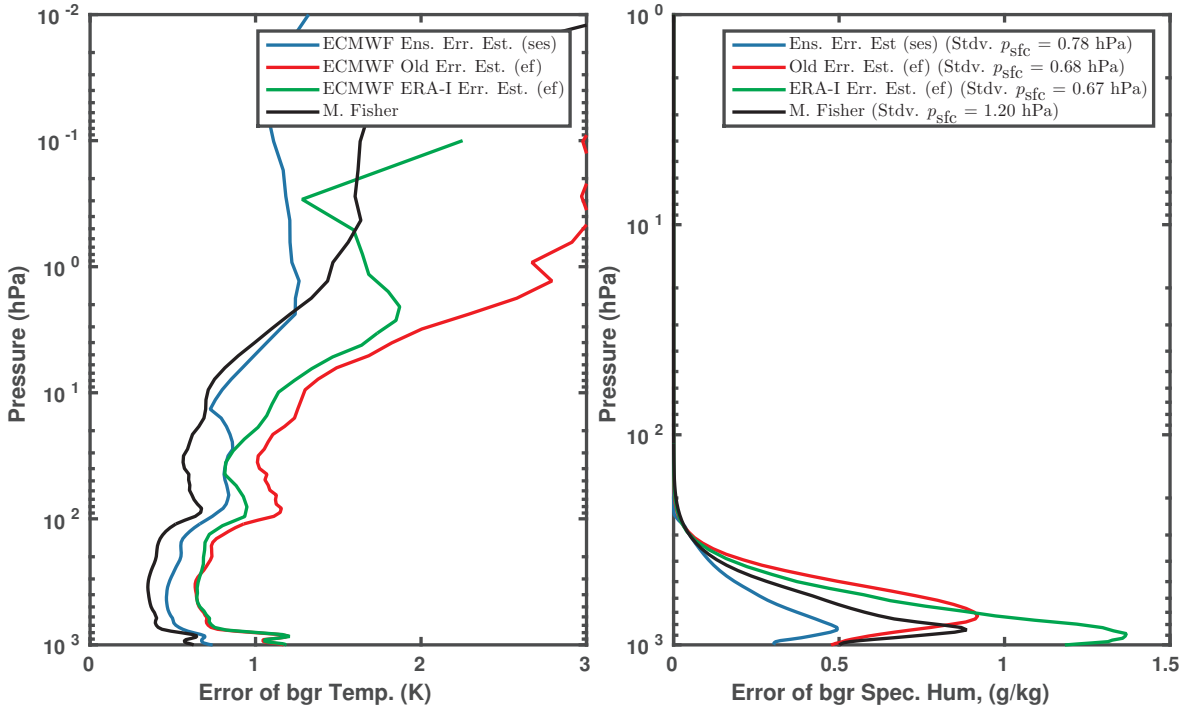


Figure 6.6: *Temperature background error profile (left) and specific humidity background uncertainty profile (right) as in Figure 6.5. Here with the uncertainty profiles for RE1 plotted in green. ECMWF has verified that the specific humidity uncertainty estimates are reasonable (S. Healy, personal communication.). The other error profiles are kept for comparison.*

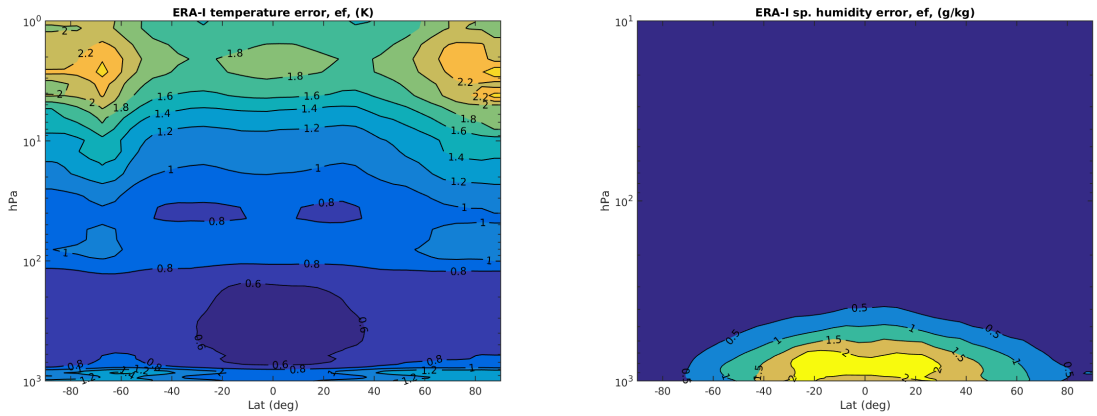


Figure 6.7: *ERA-I Temperature background uncertainty (left) and specific humidity background uncertainty (right) resolved on 5 deg. latitude bands.*

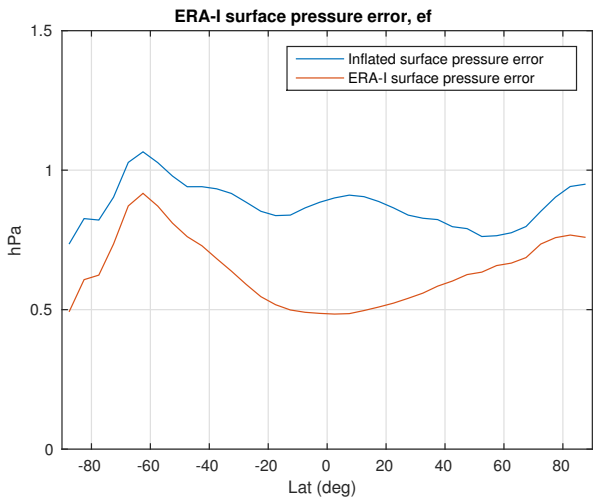


Figure 6.8: *ERA-I surface pressure background uncertainty resolved on 5 deg. latitude bands. The "inflation" is applied to include persistent bias of ERA-I forecast with respect to ERA-I analysis:* $\sigma_{\text{infl.}} = \sigma_{\text{ERA-I}} + |\overline{p_{\text{sfc,B}} - p_{\text{sfc,A}}}|$. *It was found sufficient to base the bias estimate on a couple of days from the years 2008 and 2015.*

Now, instead of the constant relative humidity method applied in NRT the latitude dependence of the specific humidity background uncertainty is represented by resolving the uncertainty model in 5 deg. latitude bands, with individual $\boldsymbol{B}$ diagonals in each band. See Figure 6.7. The error correlation matrix is kept the same for all latitude bands. In addition to this the surface pressure background uncertainty is inflated with a latitude dependent function representing a general bias in ERA-I B-A (forecast - analysis) surface pressure. See Figure 6.8.

The code and configuration for the CDR v1.0, ICDR v1.0 production is collected in GPAC v 2305. The code and configuration for the ICDR v1.1 production is collected in GPAC v 2401.

6.2.4 Special properties of the RE1 background data.

The ERA-Interim forecasts are limited to 3, 6, 9 and 12 hours. The background data are flawed due to impact of the occultations themselves between step[3] 0 and step 3 (0-3UTC and 12-15UTC), which cannot be avoided in this data set. Thus there is a systematic diurnal variability in the O-B statistics, which is to some extent seen in the S-B and S-A statistics. The impact of this "contamination" is illustrated in a series of plots below. It must be noted that the problem mixes with the variation in accuracy caused by unequal forecast times. Also note that the background error model used in 1D-Var is unchanged throughout the diurnal cycle. In Figure 6.9 (left panel) the difference in refractivty between the observation and background is plotted. Note the dip in refractivity O-B STDV following 0 UT and 12 UT. This can be a signature of contamination since information of the observations up to 3 h after forecast start may have had an imprint on the forecast, and thereby brought it close to the observation. The possible contamination is inherited in the 1D-Var solution S-B STDV

[3] In this context "step" means forecast hours. If the forecast is started from 12UTC, the "step 2" corresponds to 14UTC

which is indeed a bit lower right after 0 and 12 UT. In effect the solution is found closer to the observation in the same way it would have been, had the observation error been reduced.

In the B-A and S-A temperature plots in Figure 6.10 the forecast is closer to the analysis right after the forecast start but there is also a reduced STDV later on, after 3 UT and 15 UT, which cannot have anything directly to do with contamination. This is the effect of varying background error with forecast time. There is also a diurnal mode in the pressure statistics (Figure 6.12). But the pressure shows no sign of extraordinary reduced A-B stats following 0 UT and 12 UT. In the specific humidity case (Figure 6.11) there is a dip in the B-A STDV, following 0 UTC and 12 UTC, which is also reflected in S-A stats. In general there is a negative bias in the 1DV specific humidity at low altitude which will be addressed in the RE1 validation.

In summary there is a reduced $O-B$ standard deviation in the refractivty right after 0 UT and 12 UT which is must likely due to contamination of the background profiles. The observation is given more weight in these 3 h time intervals than elsewhere. There is no contamination effect on the mean values of any parameters.

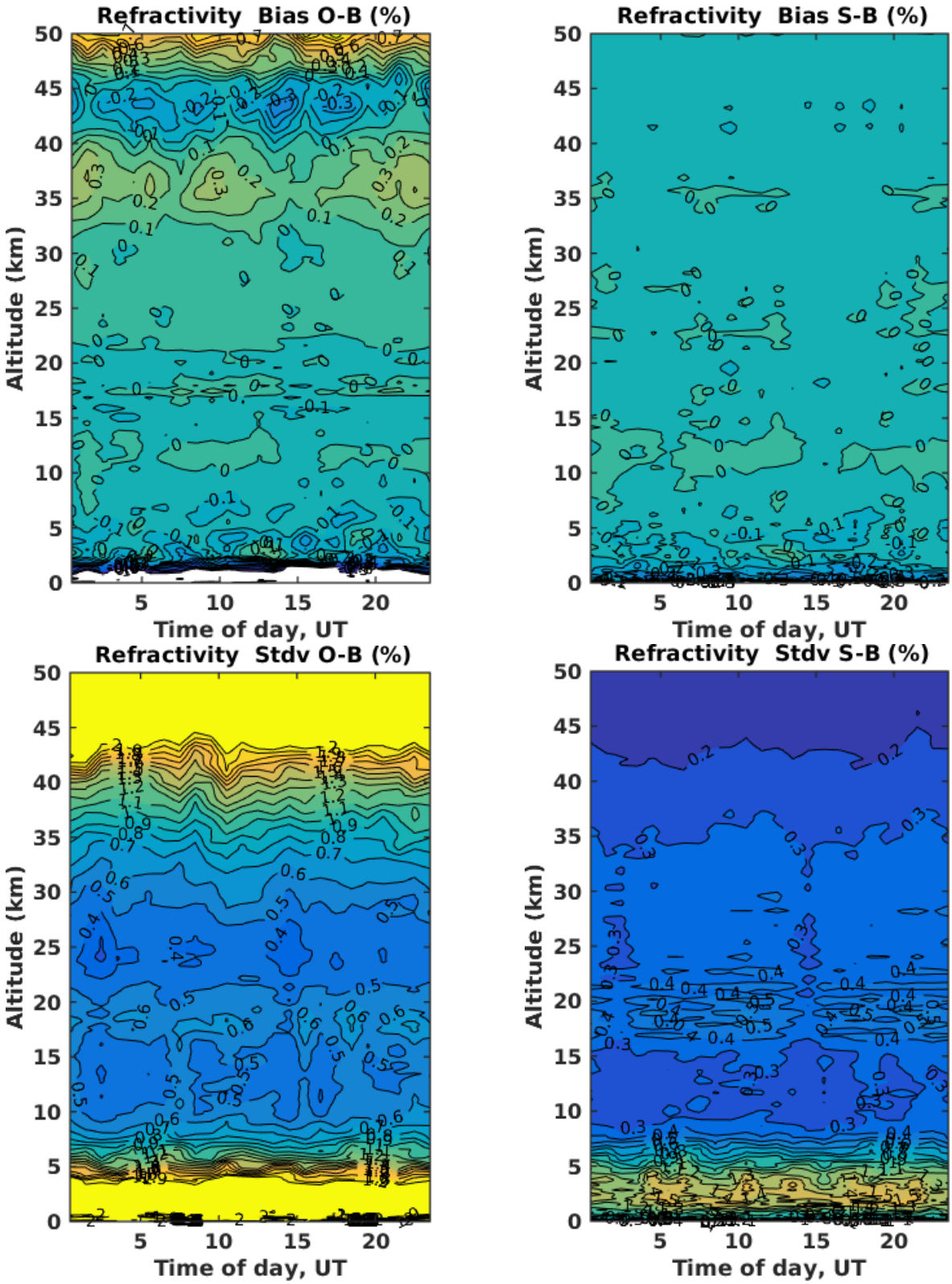


Figure 6.9: *Refractivity bias and standard deviation at different times during the diurnal cycle. 10 days og Metop and ERA-I data, February 2016. . Left: Observation Background. Right 1DV solution - Background.*

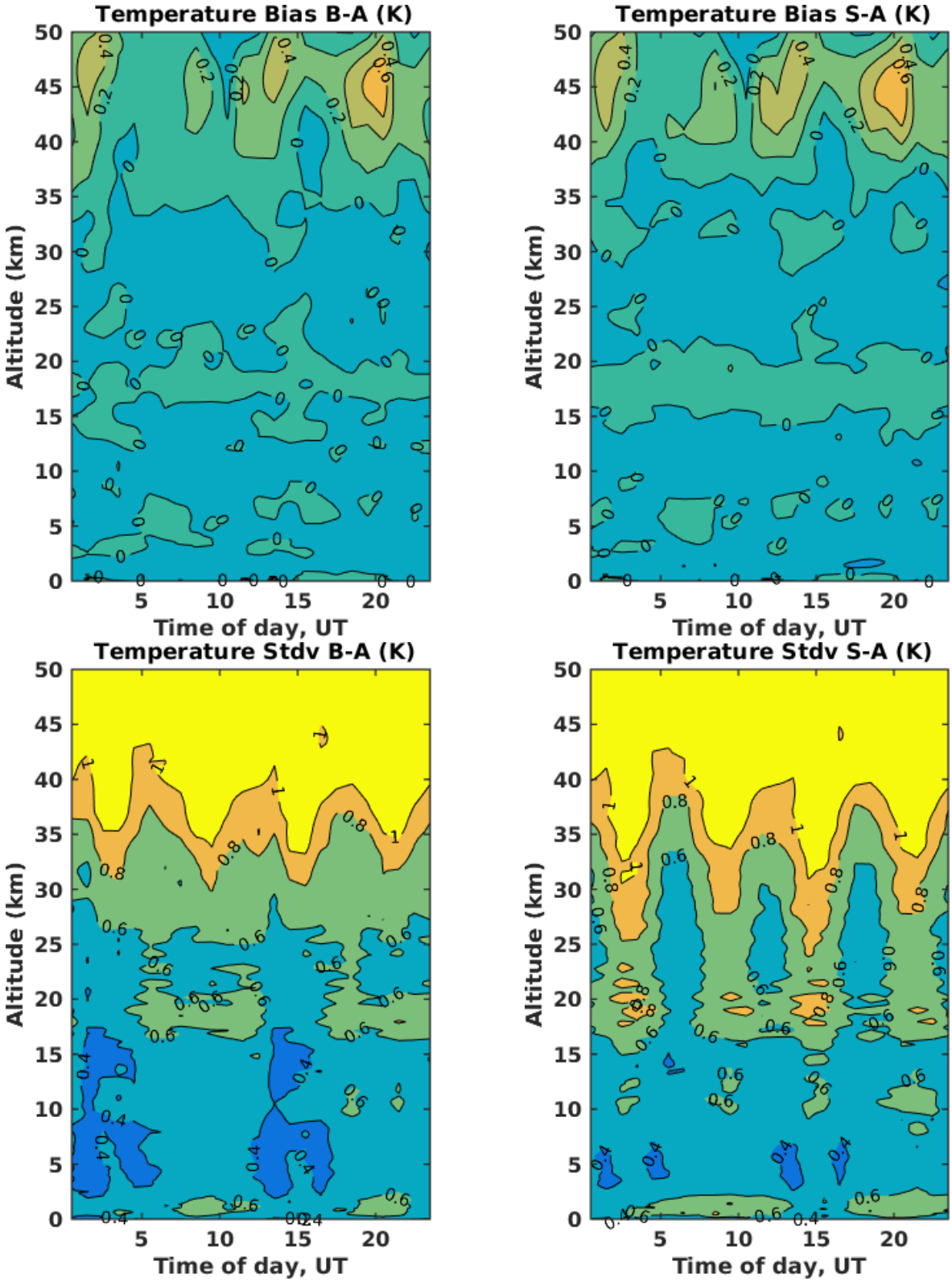


Figure 6.10: *Temperature bias and standard deviation at different times during the diurnal cycle. 10 days og Metop and ERA-I data, February 2016. Left: Background - Analysis. Right 1DV solution - Analysis.*

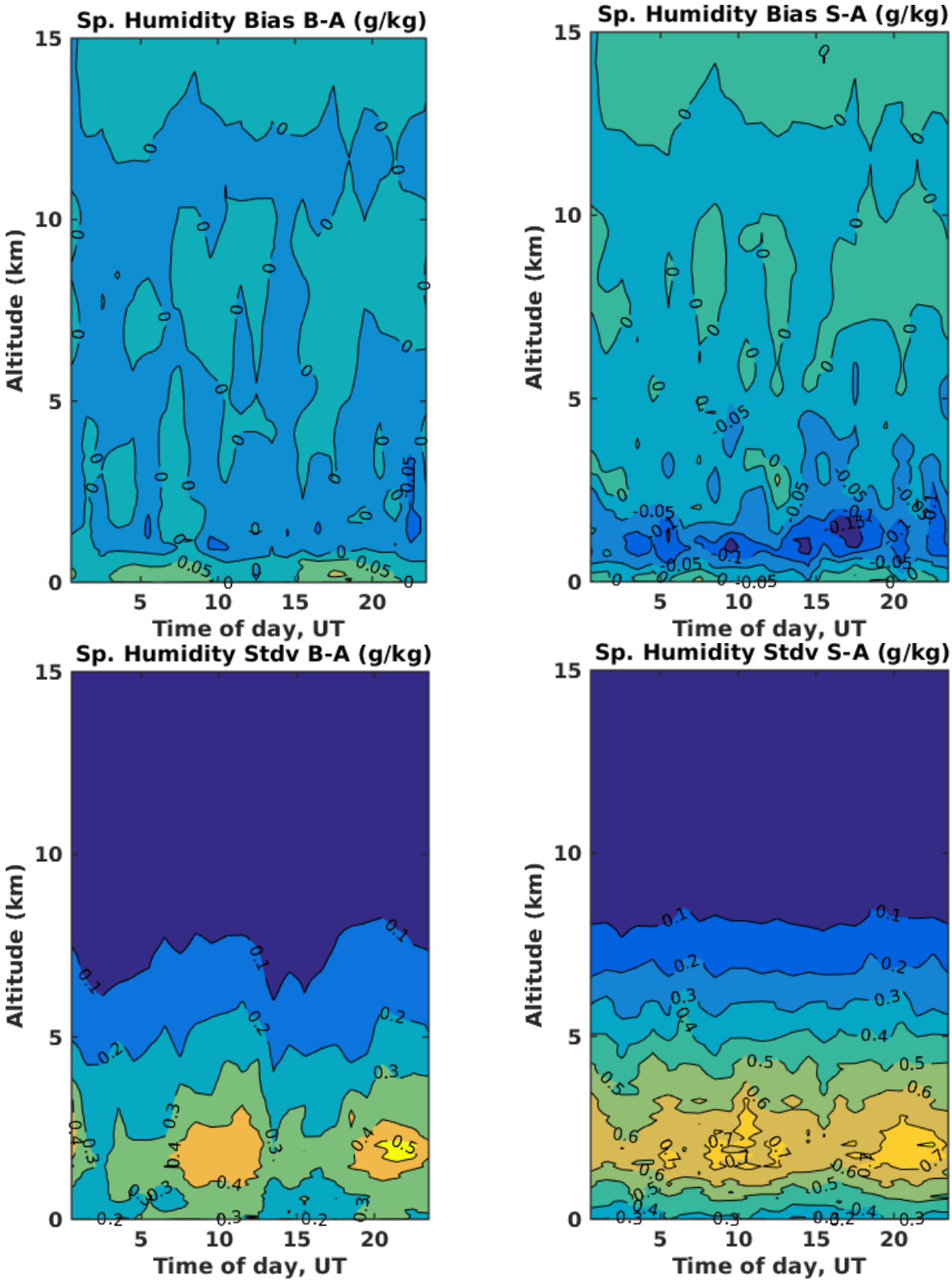


Figure 6.11: *Same as Figure 6.10, just with specific humidity.*

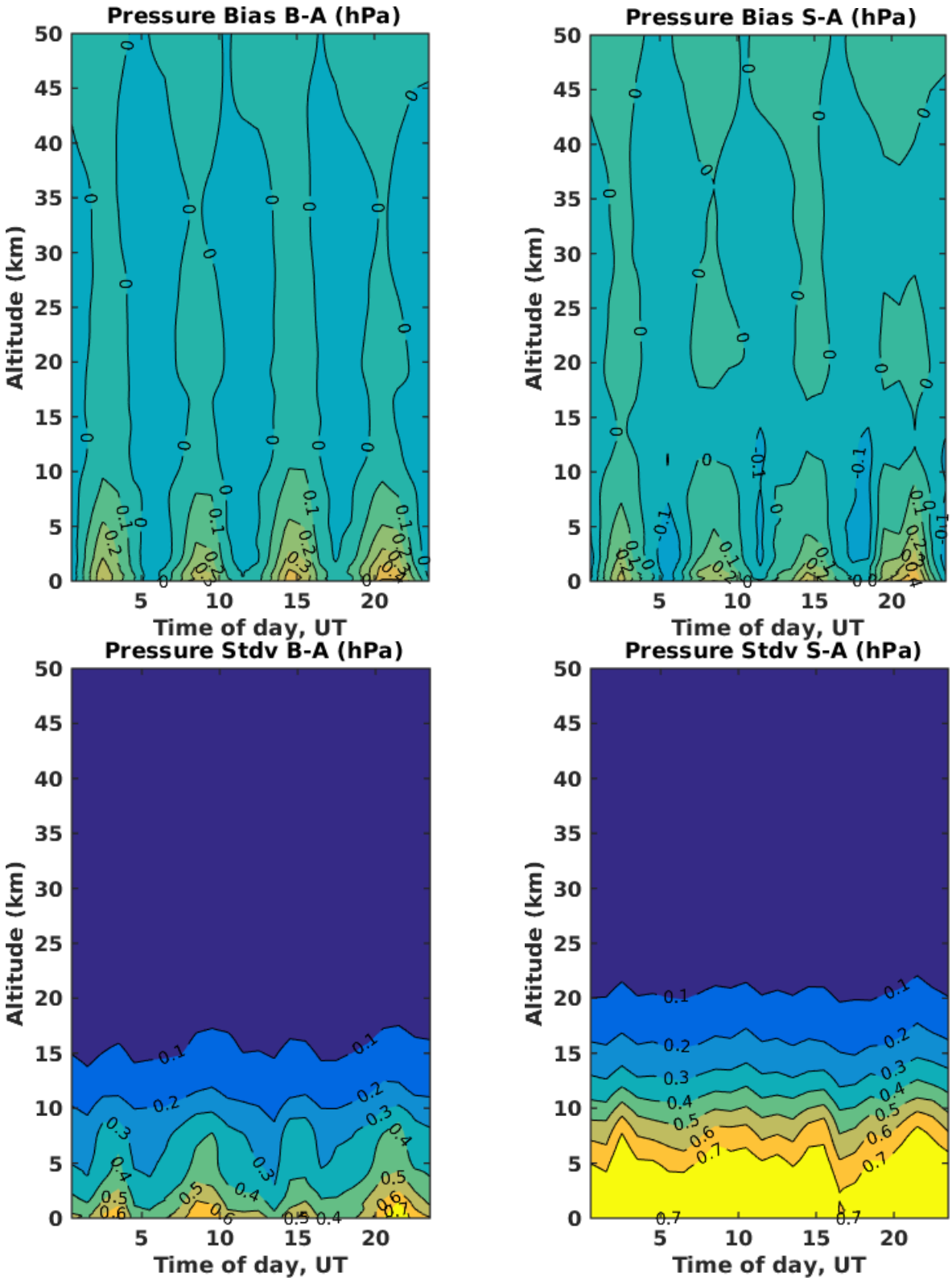


Figure 6.12: *Same as Figure 6.10, just with pressure.*

6.3 Configuration for Offline processing (1DV-3.3 and 1DV-4.2)

The 1DV Offline-v1.0 processing configuration is identical to the reprocessing (ICDR-v1.0) configuration. From 1st of August 2019 the Offline-v1.1 takes over, and unlike the ICDR v1.1, 1D-Var is performed on a full 137 level state vector after transition to ERA5. Besides that there is a change in Offline-v1.1 regarding calculation of the background tropopause, to be used in the refractivity uncertainty. In Offline v1.0 (and all previous 1D-Var products) a fallback to NCEP tropopause climatology was used every time a reasonable background tropopause was not found. This climatology fallback has been removed form Offline v1.1 and onward, such that the fallback tropopause is just set to 10,000 m in cases where no tropopause can be found, or where the found tropopause falls below 5,000 m or above 20,000 m. This change only affects the refractivity uncertainty, to be used in 1D-Var. It does not have any effect on the ROM SAF tropopause products which are calculated separately, from observations. The code and configuration for the Offline-v1.0 production is collected in GPAC v 2305. The code and configuration for the Offline-v1.1 production is collected in GPAC v 2401.

The offline products are subject to upgrades from time to time, and this will cause the offline time series to be unstable in terms of accuracy, and bias jumps are expected.

6.4 Configuration for NTC processing (1DV-4.2.2)

The 1DV NTC processing configuration is identical to the Offline-v1.1 configuration, except for a few changes:

In NTC the background fields are interpolated from ECMWF operational forecast instead of ERA5. The error covariance matrices used are identical to those used in Offline production.

Note that a few code changes and corrections have been implemented since the Offline-v1.1 products were released. In the Offline processing the forward model truncates humidity at a minimum value and interpolates humidity to observation levels on a logarithmic scale, while in NTC negative humidity values are allowed, and the humidity is interpolated linearly. Also, there has been a bug fix in the Levenberg Marquart minimizer, which may affect the 1D-Var results slightly, but does not change their statistical behavior.

The NTC 1D-Var production is based on the 1DV-v4.2.2. The code and configuration for the NTC production is collected in GPAC-v3.0.0, see table 6.4.

The NTC products are subjects to upgrades from time to time, and this will cause the NTC time series to be unstable in terms of accuracy, and bias jumps are expected.

Appendices

Appendix I. Intermediate changes to the 1D-Var code and configuration since version 1DV version 3.0

1DV Version	GPAC Version	Description	Application & date
1DV v3.0	0042	Documented in Level 2B ATBD version 2.7	Since November 2016 ROM SAF NRT has been running this version.[4]
1DV v3.1	2301	A series of software changes regarding execution control, output control, file attributes and code cleanup. No changes regarding algorithm, except that the Levenberg Marquart minimizer from this version and forth was initialized without the `-logq` and `-logp` options in order to allow negative specific humidities in the 1D-Var solution.	ROM SAF Reprocessing 1 part B, beta version, was finalized with this version in May 2017.
1DV v3.2	2302	In this version the background specific humidity covariance was produced by utilizing the ROPP latitude band resolved error covariance format. I.e., the B matrix was designed with the same global correlation structure but the diagonal was calculated as the median of the ERA-I error of first guess uncertainty (a 3 hours forecast time).	ROM SAF Reprocessing 1, part A, beta version, was finalized with this configuration in September 2017, including only Metop A/B.
1DV v3.3	2305	Documented in Level 2B ATBD version 2.9. Implemented Simmons et al [RD.30] eq. 3.18.	Applied in final version of RE1. December 2017
1DV v3.4	2401	Identical to 1DV v3.3, except for code to thin ERA5 to 60 levels.	Applied in ICDR-v1.1 starting from 1/8/2019. Implemented February 2020
1DV v4.0 (beta)		Code based completely on ROPP. Missing refractivity data levels not removed in ate. Reproduces v3.3 except that there is no fall back tropopause climatology.	This version was delivered to WEGC in April 2018 for VS35
1DV v4.1	2400	Inserted metadata. Buggy missing refractivity levels	July 2019. Interpolated 137 level B matrix based on CDR1 B matrix.

Continued on next page ...

[4] In the NRT validation report [RD.21] 1DV version 3.1 is referenced. This is not entirely correct; the `1DV_refrac` code is version 3.1, but the configuration file is version 3.0. This implies (intentional) use of the `-logq` and `-logp` options in NRT

1DV Version	**GPAC Version**	**Description**	**Application & date**
1DV v4.2	2401	Cleaned missing values in refrac_sigma	Prepared for offline poduction with 137 Level ERA5 background. February 2020
1DV v4.2.1		Upgrade of Offline processing with a few code fixes including new Lev. Marquardt, removal of hardcoded humidity boundaries in forward model and removal seg-fault in case of missing refractivity.	Prepared Offline upgrade for PPF 5.0
1DV v4.2.2	3.0.0	Using ECMWF operational (OD) background	Sentinel-6 NTC

Appendix II. Current ROM SAF 1DV Software

This appendix describes the 1DV software used in the current ROM SAF level 2B version. The current 1DV software version is 4.2. It is based on the ROPP-9.0 package with quite a few modifications collected in dmi_trunk9.0. ROPP is intended as a general-purpose programming package for the handling and processing of RO data. Complete documentation for the ROPP package is found in Refs. [RD.18, RD.19, RD.3, RD.1, RD.2].

Processing Routines

The 1D-Var processing in 1DV happens in 3 steps:

1. The input observation refractivity profile is thinned to 247 levels using the standard ROPP standalone routine ropp2ropp together with the thinning specification file `ropp_thin_eum-247.dat`. For reference see [RD.23].

2. The thinned observation profile is passed to the routine `ropp_1dvar_add_refrac_error` (source code: `ropp_1dvar_add_refrac_error.f90`). This routine constructs refractivity errors as described in section 3.7.1 and writes an observation error correlation file for each observation using the exponential expression given in equation (3.7). In order to determine the location of the tropopause, `ropp_1dvar_add_refrac_error` uses the background profile.

3. The 1D-Var processing is performed by the routine ropp_1dvar_refrac (source code: `ropp_1dvar_refrac.f90`) which takes as input the observation profile and observation error correlation file written by `ropp_1dvar_add_obs_error`, the background profile, the background error covariance file and the configuration file `romsaf_ecmwf_refrac_1dvar.cf` (see Appendix III).

Settings and Configuration

Most settings are specified in the configuration file `romsaf_ecmwf_refrac_1dvar.cf` (see Appendix III), but a few have to be set in the command line calls:

`ropp_1dvar_add_obs_error` is called with the `-Omod 'TP'` option. This option selects the refractivity error to depend on tropopause altitude as described in section 3.7.1.

The `ropp_1dvar_refrac` routine is invoked with the `-new_op, -comp, -d, - -no-ranchk` options.

Appendix III. The configuration file romsaf_ecmwf_refrac_1dvar.cf(v4.2)

```
# $Id: $

#****c* Configuration Files/romsaf_ecmwf_refrac_1dvar.cf *
#
# NAME
#    romsaf_ecmwf_refrac_1dvar.cf - Configuration file for ROM SAF operational processing
#
# SYNOPSIS
#    <1DVar_program> ... -c romsaf_ecmwf_refrac_1dvar.cf ...
#
# DESCRIPTION
#    This file reflects the configuration for the 1D-Var operational
#    processing at the ROM SAF using ECMWF background data and refractivity
#
# NOTES
#    Based on the default 1DVAR configuration file as distributed with
#    the ROPP software.
#
# EXAMPLE
#
#
# SEE ALSO
#
#
# REFERENCES
#
#
# AUTHOR
#   DMI, Copenhagen, DK, Met Office, Exeter, UK.
#   Any comments on this software should be given via the ROM SAF
#   Helpdesk at http://www.romsaf.org
#
# COPYRIGHT
#   (c) EUMETSAT. All rights reserved.
#   For further details please refer to the file COPYRIGHT
#   which you should have received as part of this distribution.
#
#****

#-------------------------------------------------------------------------------
# 1. Input and output files
#-------------------------------------------------------------------------------

# The names of input and output files can also be specified through command line
# arguments; names mspecified through the command line will always overwrite
# configuration file settings. The command line arguments corresponding to
# configuration options are given in round brackets after the detailed comments
# below.

# 1.1 Background
# --------------

bg_input_file  = ropp_bg.nc        # Background profile                          (-b, -bg, --bg)
bg_corr_file   = ropp_bg_corr.nc   # Background error correlations / covariances (--bg-corr)
```

```
# 1.2 Observations
# ----------------

obs_input_file = ropp_obs.nc       # Observation profile                          (-y, -obs, --obs)
obs_corr_file  = ropp_obs_corr.nc  # Observation error correlations / covariances (--obs-corr)

# 1.3 Output
# ----------

output_file    = ropp_out.nc       # Retrieved profile                            (-o)


#-------------------------------------------------------------------------------
# 2. Error covariance models
#-------------------------------------------------------------------------------

# 2.1 Background error covariances
# --------------------------------
#
# Background error covariances can be constructed using the following methods:
#
#   FSFC   Fixed Sigmas, Fixed correlations: Both error correlations and
#            and error standard deviations are read from a background
#            error correlation file. The error correlation file must
#            contain both the error correlation matrix as well as the
#            standard deviations (errors) for all background state vector
#            elements.
#
#   VSFC   Variable Sigmas, Fixed Correlations: Error correlations are read
#            from an error correlation file, while per profile error
#            estimates as contained in the background data file are used.
#            In this case, the error correlation / covariance data files
#            only require to contain the error correlations.
#
#   RSFC   Relative Sigmas, Fixed Correlations: Relative specific humidity (q)
#            and relative surface pressure (p*) errors, and (absolute)
#            temperature (T) errors, are read from a background error
#            correlation file, as are the error correlations.
#            (All fields must come from the same file.)  The relative q (p*)
#            errors are multiplied by the profile q (p*) values to give the
#            profile errors.  RSFC is therefore a hybrid of FSFC and VSFC.
#
# Note that error correlation files may contain latitudinally binned error
# correlations and standard deviations, allowing to have latitudinally varying
# error correlation structures and standard deviations even in the FCFS scenario.
#
# Specifying (and providing) a properly formatted error correlation / covariance
# file for the background data is mandatory for the ROPP 1DVars.

bg_covar_method  = FSFC   # Fixed Sigmas, Fixed Correlations

# 2.2 Observation error covariances
# ---------------------------------
#
# Bending angle and refractivity error covariances can be constructed using the
# following methods:
```

```
#
#    FSFC   Fixed Sigmas, Fixed correlations: Both error correlations and
#             and error standard deviations are read from an observation
#             error correlation file. The error correlation file must
#             contain both the error correlation matrix as well as the
#             standard deviations (errors) for all observation vector
#             elements.
#
#    VSDC   Variable Sigmas, Diagonal Correlations: A diagonal error correlation
#             structure (i.e., no error correlations) is assumed, while per
#             profile error estimates as contained in the observation data file
#             are used. In this case, no error correlation / covariance data
#             file is required.
#
#    VSFC   Variable Sigmas, Fixed Correlations: Error correlations are read
#             from an error correlation file, while per profile error
#             estimates as contained in the observation data file are used.
#             In this case, the error correlation / covariance data files
#             only require to contain the error correlations.
#
# Note that error correlation files may contain latitudinally binned error
# correlations and standard deviations, allowing to have latitudinally varying
# error correlation structures and standard deviations in the FSFC  and VSFC
# scenarios.
#
# In contrast to the background data, observations do not require an error
# correlation / covariance file in scenario VSDC, if the input data already
# contains error estimates.

obs_covar_method = FSFC   # Variable Sigmas, Fixed Correlations
# The ropp_1dvar_covar_refrac put sigma^2 = 0.3% on missing values with VSFC. The FSFC
# option is simpler.

#-------------------------------------------------------------------------------
# 3. Quality control
#-------------------------------------------------------------------------------

# 3.1 Valid observation height range
# ----------------------------------
#
# A valid height range can be specified, outside of which observations are
# not used in the 1dVar analysis.

min_1dvar_height = 0.0   # minimum observation height (km)

max_1dvar_height = 60.0  # maximum observation height (km)

# 3.2 Generic quality control
# ---------------------------
#
# Generic quality control checks for the obviouos things - whether data values
# are within reasonable bounds (specified below as minimum and maximum values),
# and if the input data is roughly consistent.
#
# In addition, co-location of background and observations is checked via the
# great-circle and temporal distance betwen the nominal locations of both. For
# testing purposes, and in case the 1DVar is used with climatological data as
```

```
# background which is not properly co-located or time stamped, this particular
# check can be disabled.

genqc_colocation_apply     = .false.        # Apply colocation checks?

genqc_max_distance         = 300.0          # Maximum obs vs. bg great circle distance (km)
genqc_max_time_sep         = 5400.0         # Maximum obs vs. bg temporal seperation (sec)

genqc_min_temperature      = 150.0          # K
genqc_max_temperature      = 350.0          # K

genqc_min_spec_humidity    =  0.0           # g/kg
genqc_max_spec_humidity    = 50.0           # g/kg

genqc_min_impact           = 6.2e6          # m
genqc_max_impact           = 6.6e6          # m

genqc_min_bangle           = -1.0e-4        # rad
genqc_max_bangle           =  0.1           # rad

genqc_min_geop_refrac      = -1.0e3         # m
genqc_max_geop_refrac      = 1.e5           # m

genqc_min_refractivity     = 0.0            # N-units
genqc_max_refractivity     = 500.0          # N-units


# 3.3 Background quality control
# ------------------------------
#
# Background quality control, if applied, rejects observation data points which
# deviate by more than <bgqc_reject_factor> * <expected error>, where the
# <expected error> is calculated from both the assumed observation and forward
# modelled background errors. If the number of rejected points exceeds
# <bgqc_reject_max_percent>, the entire profile is considered to be of poor
# quality, and is not further processed.

bgqc_apply                 = .true.         # Apply background quality control?

bgqc_reject_factor         = 10.0           # Data rejected if O-B > factor * sigma
bgqc_reject_max_percent    = 50.0           # Maximum percentage of data rejected


# 3.4 Probability of Gross Error
# ------------------------------
#
# Probability of Gross Error, as a diagnostic quantity, is always calculated
# as part of the standard processing (from the O - B differences). It can also
# be used for quality control, by calculating weights based on the value of
# the gross error probability which are applied to the observations.

pge_apply                  = .false.        # Apply PGE for quality control?

pge_fg                     = 0.001          # First guess PGE
pge_d                      = 10.0           # Width of gross error plateau
```

```
#-------------------------------------------------------------------------------
# 4. Preconditioning, convergence checks and minimiser
#-------------------------------------------------------------------------------

# 4.1 Preconditioning
# --------------------
#
# Preconditioning accelerates the minimisation significantly and his highly
# recommended; it should only be disabled for testing purposes.

use_precond                = .true.


# 4.2 Convergence checks
# -----------------------
#
# Apart from the minimiser's own convergence check (which is based on the change
# of the gradient size), two more convergence checks are implemented: The
# minimisation is assumed to have converged if any of the following conditions
#
#  - change of state vector < <conv_check_max_delta_state> * <background error>
#  - change of cost function < <conv_check_max_delta_J>
#
# is achieved for at least <conv_check_apply> consecutive iterations / calls of
# the cost function evaluation.
#
# Note that the convergence criteria for the state vector changes are specified
# as a fraction of the background error.

conv_check_apply           = .true.         # Apply additional convergence checks?

conv_check_n_previous      = 2              # Minimum number of iterations required
conv_check_max_delta_state = 0.1            # State vector must change less than this * bg error
conv_check_max_delta_J     = 0.1            # Cost function must change less than this


# 4.3 minROPP settings
# -------------------

minropp_method               = LEVMARQ        # minimisation method
                                              # MINROPP (default) or LEVMARQ
minropp_log_file             = screen         # 'screen' for output on the screen, file name otherw
minropp_impres               =    0
minropp_n_iter               =  1500
minropp_mode                 =    0
minropp_n_updates            =   50
minropp_eps_grad             = 1.0e-8
minropp_dx_min               = 1.0e-16


#-------------------------------------------------------------------------------
# 5. Additional output
#-------------------------------------------------------------------------------
#
# If <extended_1dvar_diag> is .true., the 1DVar will add additional diagnostic
# output to the retrieval data file. At present, the following diagnostics are
# available:
```

```
#
#  - O-B differences
#  - O-A differences
#  - PGE

extended_1dvar_diag     = .true.
# Reenforced by -d option in ropp_1dvar_refrac call

#-------------------------------------------------------------------------------
# 6. Log(pressure) and Log(humidity) options
#-------------------------------------------------------------------------------
#
# If <use_logp> is .true., the 1DVar will perform the minimisation using log(p)
# rather than absolute pressure in the state vector
# If <use_logq> is .true., the 1DVar will perform the minimisation using log(q)
# rather than absolute pressure in the state vector

use_logp = .false.
use_logq = .false.

#-------------------------------------------------------------------------------
# 7. Seasonal observation scaling options
#-------------------------------------------------------------------------------
#
# These options allow the observation errors (the standard deviation values) to
# be scaled according to the season. This is a sinusoidal correction of the form:
# new_error=old_error*(1+<season_offset>+<season_amp>*COS(2pi*(season+<season_phase>))).
# 'season' is calculated in the code and takes a value between 0 and 1 (the start
#     and end of the calendar year respectively)
# <season_amp>     is the amplitude of the sinusoidal scaling factor.
# <season_offset>  is a constant offset applied to the observation errors, on
#    which the sinusoidal factor is added.
# <season_phase>   gives control over the 'phase' of the sinusoid, i.e. a value
#    of 0.1 will shift the maximum of the sinusoid back one tenth of a year.
#    This value should be between -1 and 1.
#
# Take care using this functionality as it is possible to produce negative
# standard deviation values. In this case a warning is produced and the program
# reverts to the unscaled sigma values.
# A recommended approach is to input observation errors for the season which has
# the smallest errors and provide a positive value for <season_offset> which is
# equal to <season_amp>. This will ensure that the seasonal adjustment will
# always result in positive variances.

season_amp    = 0.0
season_offset = 0.0
season_phase  = 0.0
```

Appendix IV. Correction to full level pressure interpolation

For reference we summarize the impact of implementing the Simmons et al. ([RD.30]) method instead of simple linear average of closest half levels for calculating full level pressures.

The "usual" full level pressure interpolation referred to is

$$p_k = \frac{1}{2}\left(p_{k-\frac{1}{2}} + p_{k+\frac{1}{2}}\right) \tag{6.4}$$

where $k \mp \frac{1}{2}$ is short notation for the half levels above/below level k.

A more accurate approximation of the full level pressure, which is consistent with the full level geopotential height, is given in [RD.30] eq. 3.18:

$$p_k = \exp\left[\frac{1}{\Delta p_k}(p_{k+1/2}\ln p_{k+1/2} - p_{k-1/2}\ln p_{k-1/2}) - C\right], \tag{6.5}$$

where $C = 1$ at all levels except at the top level where $C_1 = \ln 2$.

Adoptation of equation 6.5 has considerable impact on the calculated background and analysis profiles in from ERA-I and on the resulting solution when applied in 1D-Var. The pressure corrections for two different interpolation methods are plotted in Figure 6.13. For future reference we included the other suggested form ([RD.30] eq. 3.17).

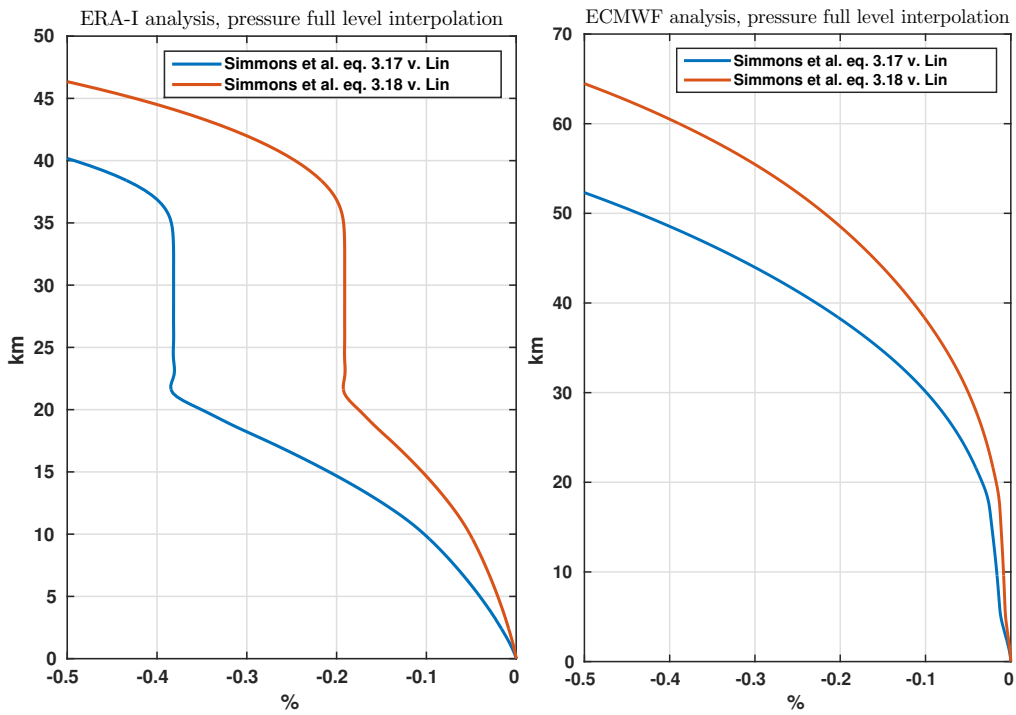


Figure 6.13: *Relative pressure interpolation biases (note the different altitude scales); Left: effect of Simmons eq. 3.17 / eq. 3.18 on ERA-I wrt. ERA-I "normal reference" (eq. 6.4); Right: Effect of Simmons eq. 3.17 / eq. 3.18 on ECMWF wrt. ECMWF "normal reference" (eq. 6.4).*

The effect of applying Simmons eq. 3.18 is basically a negative shift of all pressures, as seen in Figure 6.13 (red curves). For ERA-I the pressure shift is of considerable magnitude. A pressure correction of 0.2 % (eg. at 30 km) corresponds to a lift of about 20 m. For ECMWF(OPER) the correction has less impact, though in the upper stratosphere it could possibly have effects on eg. O-B statistics. This is not being pursued further in here. ERA5

has inherited the level structure of ECMWF (OPER), so it can be assumed that the impact on ERA5 will be similar to what is shown on the right axis of Figure 6.13.

The effect of applying the form in eq. 6.5 is shown for various properties in the following figures. Starting with the refractivity in Figures 6.14 and 6.15 it is noted that the adjustment largely removes the negative O-B bias between 10 and 30 km, while it introduces a positive bias around 35 km. The solution pressure biases are generally removed (figure 6.16). Regarding the temperature solution (figure 6.16) it is reassuring to note that the negative S-B (solution - background) bias in the lower troposphere which has been hard to rationalize otherwise has now been greatly reduced. The error effected 1D-Var in the following way: The positive pressure bias in the background field caused 1D-Var to seek a solution with lower pressure than the background. This caused the geopotential height for a given model level to be lower in the solution as well, and since the temperature background error is relatively small in the troposphere, the temperature for a given level was not allowed to change much. This means that the solution temperature was still essential equal to the background temperature, but it was evaluated at a lower geopotential height, i.e against a higher background temperature, and this caused the apparent negative temperature S-B bias. Likewise, the specific humidity (solution-analysis) bias has also been reduced (figure 6.18). There is still a residual negative bias in humidity which is expected since the measured refractivity is know to be negatively biased in the boundary layer.

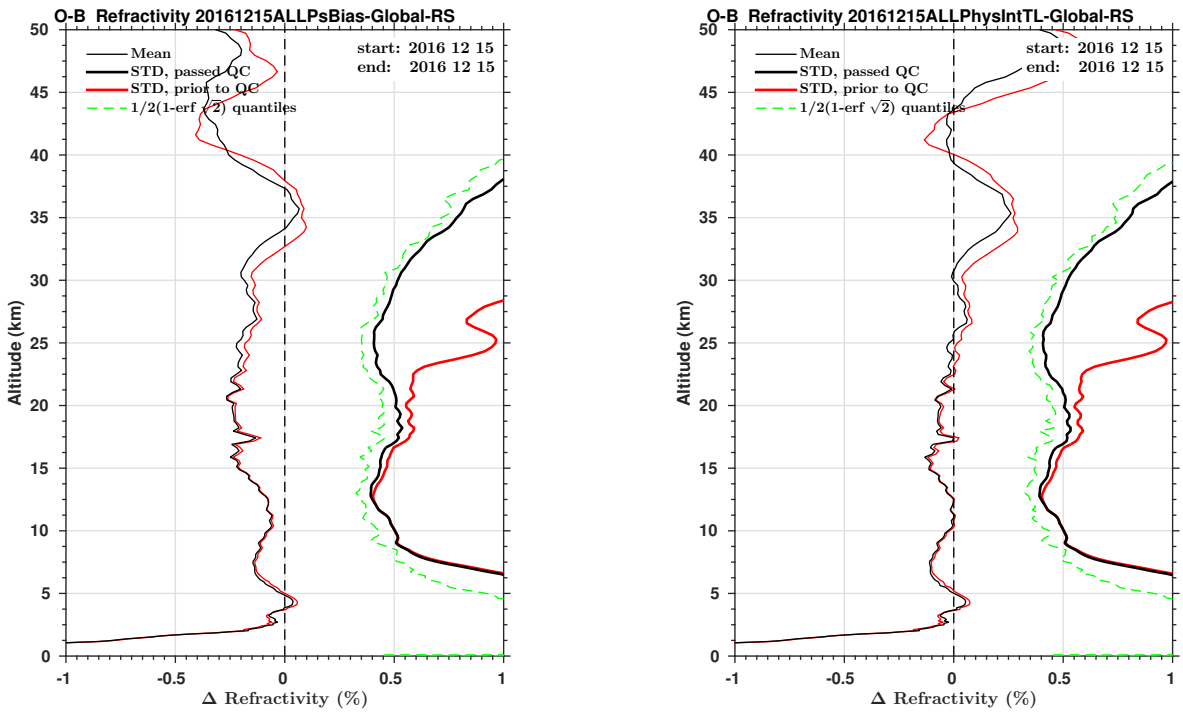


Figure 6.14: *O-B, Refractivity v. ERA-I background. Thin curves: mean, thick curves: STDV, red: without QC, black with QC, dashed green: quantile corresponding to STDV in Gaussian distribution. Linear (l), Simmons et. al eq. 3.18 (r)*

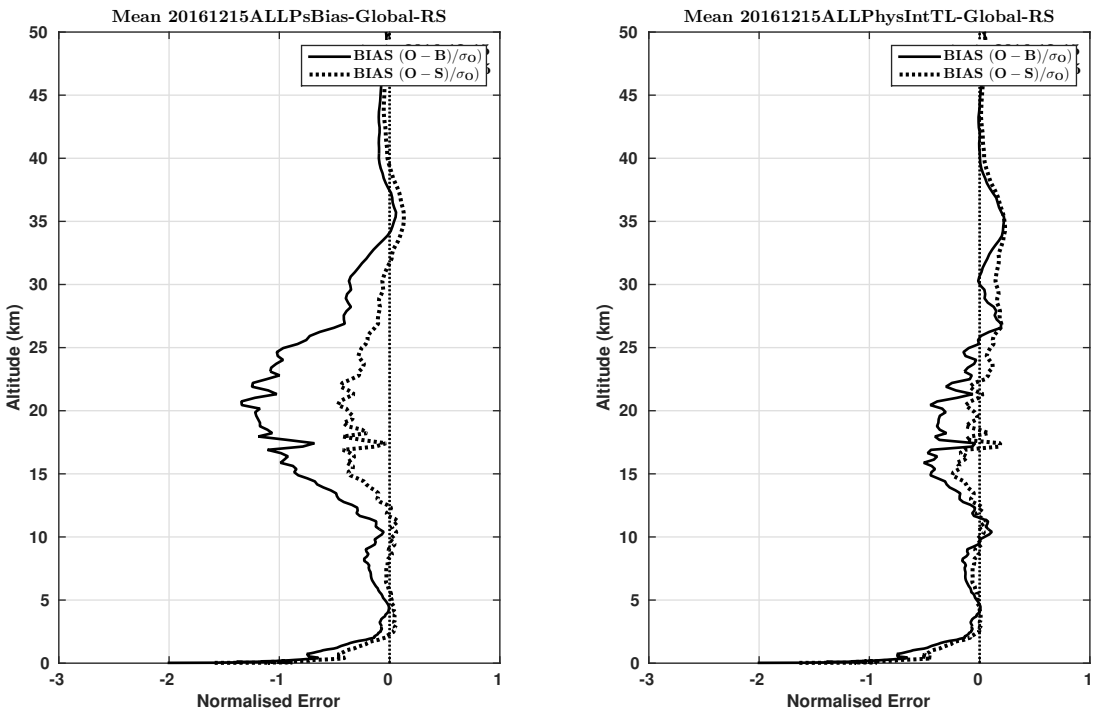


Figure 6.15: *$(O-B)/\sigma_B$, Refractivity v. ERA-I background. Refractivity v. ERA-I background. Thin curves: mean, thick curves: STDV, red: without QC, black with QC, dashed green: quantile corresponding to STDV in Gaussian distribution. Linear (l), Simmons et. al eq. 3.18 (r)*

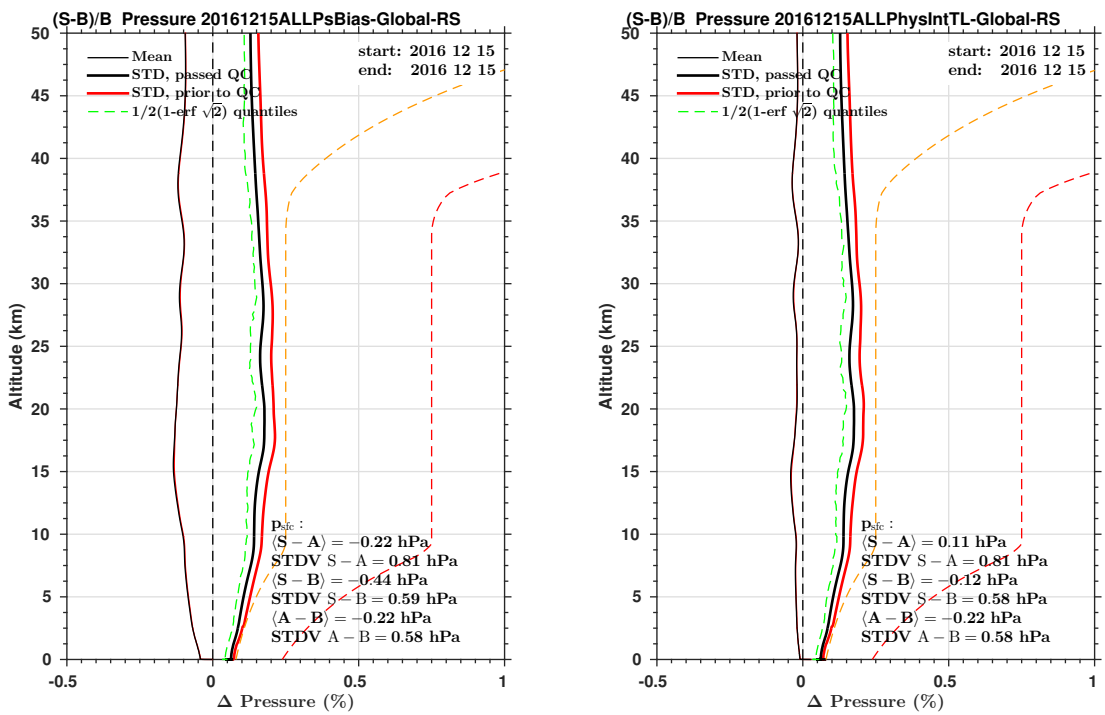


Figure 6.16: *$(S-B)/B$ Pressure. Refractivity v. ERA-I background. Thin curves: mean, thick curves: STDV, red: without QC, black with QC, dashed green: quantile corresponding to STDV in Gaussian distribution. Linear (l), Simmons et. al eq. 3.18 (r)*

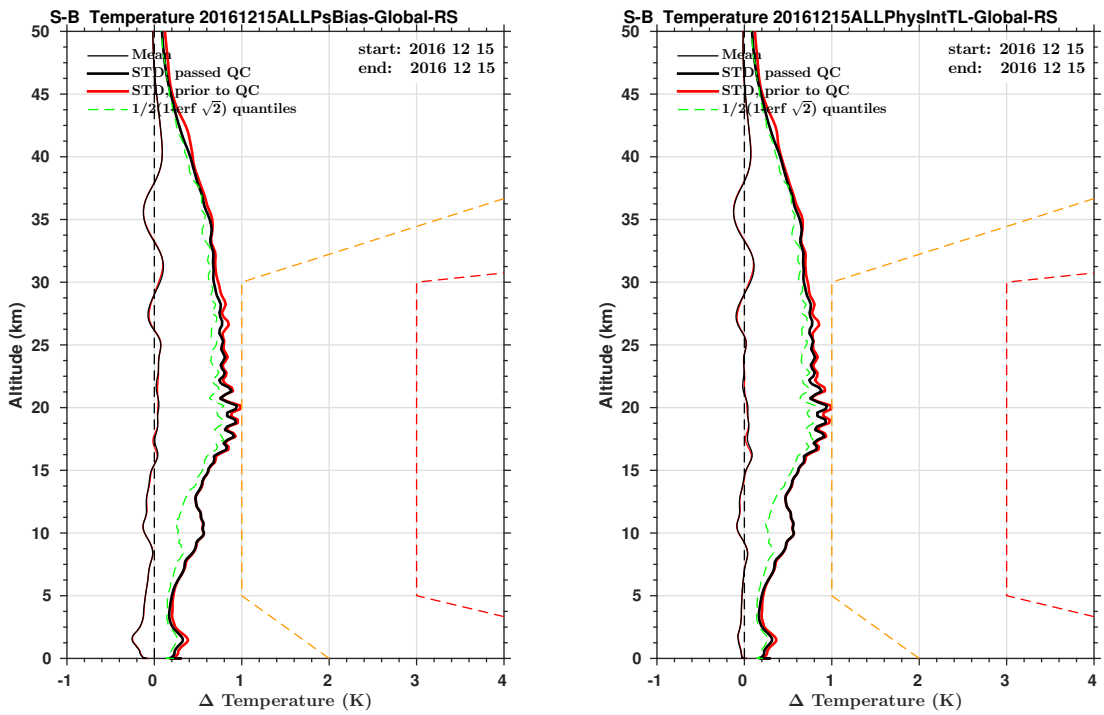


Figure 6.17: *(S − B) Temperature. Refractivity v. ERA-I background. Thin curves: mean, thick curves: STDV, red: without QC, black with QC, dashed green: quantile corresponding to STDV in Gaussian distribution. Linear (l), Simmons et. al eq. 3.18 (r).*

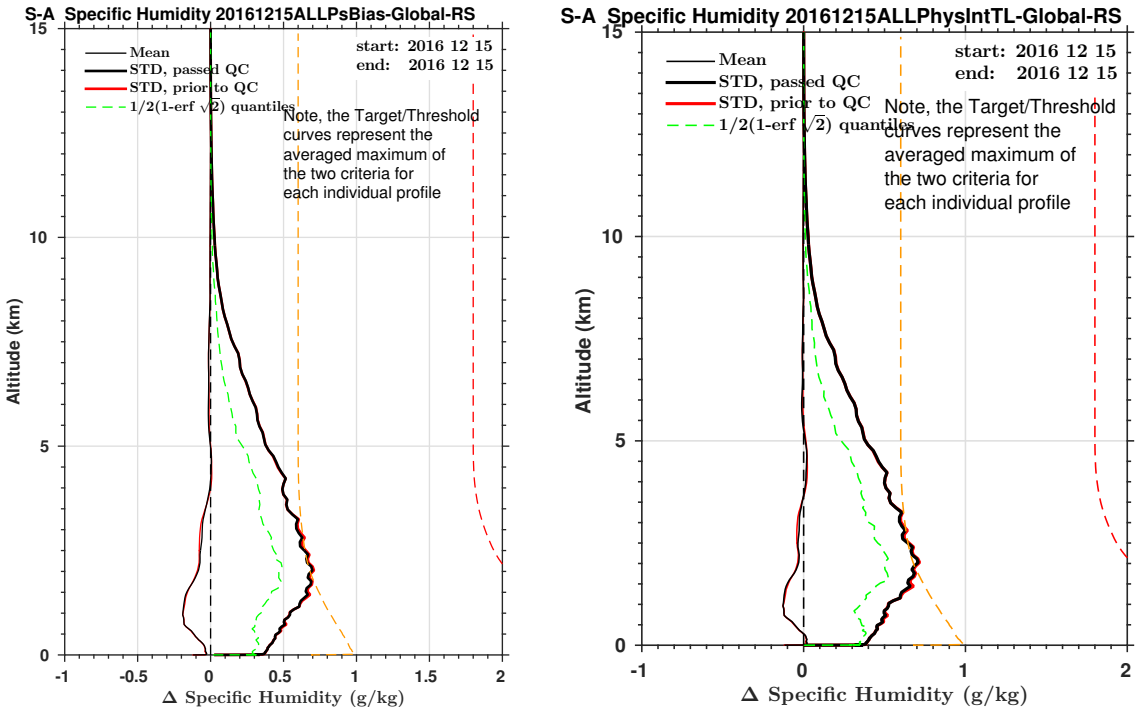


Figure 6.18: *(S − B) Specific Humidity. Refractivity v. ERA-I background. Thin curves: mean, thick curves: STDV, red: without QC, black with QC, dashed green: quantile corresponding to STDV in Gaussian distribution. Linear (l), Simmons et. al eq. 3.18 (r).*