

***MTG-IRS L2 ATBD***

Doc.No. : EUM/RSP/TEN/17/935387  
Issue : v1F e-signed  
Date : 16 September 2021  
WBS/DBS :

EUMETSAT  
Eumetsat-Allee 1, D-64295 Darmstadt, Germany  
Tel: +49 6151 807-7  
Fax: +49 6151 807 555  
<http://www.eumetsat.int>

## Change Record

<b>Version</b>	<b>Date</b>	<b>DCR* No. if applicable</b>	<b>Description of Changes</b>
V1			Initial version of the IASI L2 Day-1 baseline based on IASI operational heritage
V1A			Editorial updates after internal review
V1B			Updates signature table
V1C			<p>Updates after IRS MAG review – May 2018</p> <p>Mostly editorial, clarifications and typo fixes</p> <p>Main changes:</p> <ul style="list-style-type: none"> <li>• Added generation of instability indices and corresponding references (§3.10.2, §3.10.3, §4.13, §4.15)</li> <li>• Added generation of error estimate profile, from PWLR3</li> <li>• Updated open issues, for Day-2</li> <li>• Regrouped L2 products, error covariance in separated product</li> </ul>
V1D			Same as v1C, all changes accepted – for publication
V1E draft			<p>28/11/2019</p> <p>Typo fixes</p> <ul style="list-style-type: none"> <li>- §3.10.3.6.1 missing factor <math>q_i</math> in eq.65</li> <li>- Sign of equation 71. Condition in eq.72</li> <li>- Including CO<sub>2</sub> retrieval in PWLR3 retrieval (§3.6.2.5)</li> </ul> <p>Removed Appendixes D, E - intended to include supporting results, which have been presented to MAG meanwhile (EUM/RSP/VWG/17/948734).</p> <p>Formula update in the cloud retrieval (§3.7.1.2), using observation error covariance matrix instead of a vector of weights (diagonal matrix) + bias correction.</p>
V1E			Read-only for publication - 20 January 2020
V1F	16/09/2021	839	<p>Signature table update</p> <p>Editorial and minor clarifications for publication</p> <p>Further explicated the nature of the water-vapour profiles used as (specific humidity vs mixing ratio)</p> <p>§3.6.2.5 Repeat explicitly in this paragraph the nature of uncertainty estimates (profiles and vector for atmospheric parameters and surface emissivity)</p> <p>Small update to open issues</p>

\*DCR = Document Change Request

## **Table of Contents**

<b>1 INTRODUCTION .....</b>	<b>8</b>
1.1 Purpose .....	8
1.2 Scope .....	8
1.3 Applicable Documents .....	8
1.4 Reference Documents .....	8
1.5 Document Structure .....	8
<b>2 AN INTRODUCTION TO THE MTG-IRS MISSION .....</b>	<b>10</b>
2.1 The Meteosat Third Generation Programme .....	10
2.2 MTG-IRS target applications .....	10
2.3 The satellite platform .....	11
2.4 A major innovation with strong applicable hyperspectral infrared heritage .....	13
2.5 Specificities: limitations and opportunities .....	14
2.5.1 Spectral coverage and resolution .....	15
2.5.2 No microwave instrument companion, higher sensitivity to clouds .....	15
2.5.3 Viewing geometry, from Nadir to quasi-limb viewing .....	16
2.5.4 Spatial resolution and coverage.....	18
2.5.5 Temporal sampling .....	19
<b>3 DAY-1 PROCESSING SEQUENCE .....</b>	<b>21</b>
3.1 Common HSIR L2 approach.....	21
3.2 Processor inputs .....	21
3.3 Output end-products .....	22
3.4 Processing sequence overview.....	23
3.5 Pre-processing .....	24
3.5.1 Input preparation and collocation.....	24
3.5.2 Acceptance of L1 measurements .....	24
3.5.3 Collocation of Digital elevation model and land-mask .....	24
3.5.4 Collocation of NWP forecasts .....	25
3.6 First retrieval: PWLR <sup>3</sup> , an “all-sky” 3D statistical retrieval.....	26
3.6.1 Algorithm concept and rationale .....	27
3.6.1.1 PieceWise Linear Regression: a non-linear statistical retrieval .....	27
3.6.1.2 Training with real observations.....	28
3.6.1.3 Defining the regression classes .....	29
3.6.1.4 The PWLR <sup>3</sup> , a 3D retrieval .....	30
3.6.2 Detailed description of the algorithm and configuration.....	32
3.6.2.1 Training set.....	32
3.6.2.2 Examination and correction of the surface temperature in the training set.....	34
3.6.2.3 Definition of the cloud signal parameter.....	34
3.6.2.4 Computation of quality indicators .....	35
3.6.2.5 Detailed description of the retrieval algorithm .....	37
3.6.2.6 Moving window strategy .....	41
3.6.3 Handling non-nominal situations.....	42
3.7 Cloudy scene characterisation .....	43
3.7.1 Cloud retrieval .....	43
3.7.1.1 Retrieval principle .....	43
3.7.1.2 Computing the costs with clear and cloudy assumptions .....	44
3.7.1.3 Clear/Cloudy costs assessment and scene classification.....	46
3.7.1.4 Input profiles to RTTOV simulations for cloud retrievals .....	47
3.7.1.5 Channel selection and weighting for cloud retrievals .....	47
3.7.2 Cloud phase .....	47
3.7.3 Dust detection .....	48
3.8 Second retrieval: Optimal Estimation Method.....	49
3.8.1 Overview .....	49
3.8.2 Background term of the cost function, $Jx$ .....	49
3.8.2.1 Choice of <i>a priori</i> , $xa$ .....	50
3.8.2.2 <i>A priori</i> error covariance matrix, $Sx$ .....	51

3.8.3	Observation term of the cost function, $Jy$ .....	51
3.8.3.1	Forward model, F .....	51
3.8.3.2	Observations, $y$ .....	52
3.8.3.3	Observation error covariance matrix, $Sy$ .....	55
3.8.4	Minimization of the cost function, $J = Jx + Jy$ .....	55
3.8.5	Retrieval error covariance matrix .....	56
3.9	Quality control .....	57
3.10	Post-processing .....	57
3.10.1	Geometric corrections .....	57
3.10.1.1	Reconstruction of vertical profiles .....	57
3.10.1.2	Geolocation of the clouds .....	60
3.10.1.2.1	Computation of columnar amounts .....	61
3.10.1.2.2	Computation of stability indices .....	61
3.10.1.3	Lifted Index .....	64
3.10.1.3.1	Compute $t$ .....	65
3.10.1.3.2	If $t < 0$ then : .....	65
3.10.1.3.3	else: .....	65
3.10.1.4	K-Index, KI .....	66
3.10.1.5	Layer-Precipitable Water, LPW .....	66
3.10.1.6	Maximum Buoyancy, MB .....	66
3.10.1.7	$\Delta\theta_e$ .....	67
3.10.1.8	Available Potential Energy indices, CAPE and CIN .....	67
3.10.1.8.1	Procedure to computing convective inhibition and available potential energy ..	68
3.10.1.8.2	Surface-based CAPE (SBCAPE or CAPE) and CIN .....	69
3.10.1.8.3	Mixed-Layer CAPE, MLCAPE .....	69
3.10.1.8.4	Maximum Unstable CAPE, MUCAPE .....	69
4	<b>GENERIC FUNCTIONS</b> .....	71
4.1	Vertical interpolations of atmospheric temperature and constituent profiles .....	71
4.2	Vertical integration of atmospheric constituent concentrations .....	71
4.3	Conversion from pressure to height levels .....	72
4.4	Linear interpolation .....	73
4.5	Bilinear interpolation .....	73
4.6	Euclidean norm .....	73
4.7	Dot product .....	73
4.8	Vector product .....	74
4.9	Water-vapour density definitions and relationships .....	74
4.10	Virtual temperature .....	76
4.11	Check for Super-saturation of Water Vapour .....	76
4.12	Moist air specific heat and gas constant .....	76
4.13	Atmospheric adiabatic lapse rates .....	77
4.14	Check for Super-Adiabatic Layering .....	77
4.15	Lifting condensation level (LCL) .....	78
4.16	Slanted levels geolocation .....	78
	<b>APPENDIX A ASSUMPTIONS AND OPEN ISSUES</b> .....	82
	<b>APPENDIX B REFERENCES</b> .....	84
	<b>APPENDIX C ACRONYMS</b> .....	89
	<b>APPENDIX D DRAFT PRODUCTS CONTENT TABLES</b> .....	91
D.1	IRS-2-SVP .....	91
D.2	IRS-2-COV .....	94
	<b>APPENDIX E RADIANCE BIAS CORRECTION</b> .....	96
	<b>APPENDIX F EXAMPLE OF A C++ IMPLEMENTATION OF THE SUB-PROFILE POINTS COMPUTATIONS</b> .....	101

## Table of Figures

Figure 1: MTG-IRS dwells and Local Areas Coverage .....	12
Figure 2: Quickview of IRS pixel on-ground dimension [km] .....	13
Figure 3: Comparative spectral coverage of IRS and IASI .....	15
Figure 4: Viewing geometry principles for IRS (top) and IASI (bottom) .....	17
Figure 5: IRS local zenith angle at target on Earth ( $z$ in Figure 4) as a function of angular distance from sub-satellite point ( $\alpha$ in Figure 4). ....	17
Figure 6: Areas corresponding to the maximum local zenith angle with IASI (within the blue disk), used for assimilating MSG/SEVIRI radiances in NWP models (orange) and deriving instability indices from SEVIRI (red). The purple line shows the $85^\circ$ zenith angle limit of applicability of RTTOV (nearly on disk edge in the left panel). ....	18
Figure 7: Examples of IRS Jacobians for temperature (left panel) and water-vapour (right) at different viewing angles.....	18
Figure 8: Average cloud-free pixel percentage as a function of pixel size, from [Wang et al. 2016]....	19
Figure 9: Simulation of an IRS scene using high resolution model data from Meteo-France, with realistic dwell limits (pink and blue) and IASI footprints from real observations in that area (black ellipses). ....	20
Figure 10: High level overview of IRS L2 processing steps – the optional use of NWP forecasts in the cloud and OEM retrieval steps instead of PWLR <sup>3</sup> profiles is indicated with the dash yellow box	23
Figure 11 ECMWF forecast files sequence and selection for interpolation to sensing time $t$ (red line).....	25
Figure 12: Observed slanted profile (red) and vertical 1D profile (green) as stored in the forecast data.....	26
Figure 13: Piecewise Linear Regression concept. Linear relationship is modelled in different classes of observations, here illustrated with the 3 segments. In real product processing, the clustering is based on the observations by application of k-mean techniques. ....	28
Figure 14: Linear regression (plain blue) modelling true (dashed green) function from experimental collection of input/target pairs (red dots) in an unbiased (left) and a biased (right) sample. ....	29
Figure 15: An example of PWLR clustering based on IASI observations.....	30
Figure 16: Maps of principal component scores in common (in PC #2, left) and interpixel differences (in PC #3, right) directions of 4 adjacent IASI pixels.....	31
Figure 17: Eigenvectors for IASI pixel (IFOV) quadruplets, carrying common (EV1-2) and interpixel (EV3-6) information. ....	31
Figure 18: IASI observations (OBS) minus calculated radiances, using different collocated ozone sources (purple: CAMS analysis, blue: ECMWF IFS forecasts, red: ERA-5, black: retrievals based on ERA-5).....	33
Figure 19: Standard deviation of OBS – CALC radiances computed over continental surfaces on 16/05/2016 in the first two IASI bands. The inputs to the forward computations are collocated ECMWF forecasts + static emissivity atlas (black), PWLR (blue) and PWLR <sup>3</sup> profiles and surface parameters retrievals (red). ....	35
Figure 20: Errors between true parameters (dots) and regression estimates (line) .....	36
Figure 21: PWLR <sup>3</sup> adjacent pixels grouping. Principal components of PCs in each individual IRS pixel are computed first in the 2x2-pixel quadrants (blue, green, red, yellow) and then combined again to form the inputs to the PWLR <sup>3</sup> retrieval (white box)., .....	37
Figure 22: PWLR <sup>3</sup> retrieval sequence overview .....	38
Figure 23: Stratification of Temperature statistics (dash: bias, solid: standard deviation) IASI-PWLR <sup>3</sup> vs ECMWF analysis on 04/10/2016, partitioning in 4 quality classes of equal size based on quality indicator .....	41
Figure 24: Map of the quality classes defined above.....	41
Figure 25: PWLR <sup>3</sup> moving window strategy illustrated on an 8 by 8 IRS-pixels subgrid of a dwell. Left: non-overlapping groups. Right: overlapping groups .....	42
Figure 26: Schematic view of IRS satellite zenith angle wrt local zenith (yellow).....	58
Figure 27: Observed slanted (purple) and vertical profile provided to the End-Users (green). .....	58
Figure 28: Reconstructing profiles at the vertical of a pixel with slant retrievals. ....	59
Figure 29: Apparent and actual geolocation of a cloud retrieved on a slant path (IRS pixels on ground illustrated with the black boxes) .....	60

Figure 30: Conditional instability illustrated with a real atmospheric sounding from IASI-B near Tuzla (Bosnia and Herzegovina) on 10 August 2018. ....	63
Figure 31: Viewing geometry of target and sub-profile points .....	79
Figure 32: IASI OBS-CALC statistics computed with the clear ocean cases identified in the IAVISA database at mid and low latitudes.....	97
Figure 33: Mean IASI CALC-OBS computed with ARSA sonde measurements and RTTOV .....	98
Figure 34: Histogram of the differences IASI “CALC-OBS” at $704.5 \text{ cm}^{-1}$ ( $\text{CO}_2 15\mu\text{m}$ band) with the ARSA sonde database: green= sea day, blue = sea night, purple = land day, red = land night .	98
Figure 35: Mean (top) and standard deviation (bottom) difference between LBLRTM v11.1 and v12.2 simulations with different 5190 profiles [Matricardi 2014] .....	99
Figure 36: Mean (top) and standard deviation (bottom) difference between RTTOV and LBLRTM v12.2 simulations of IASI maritime (left) and continental (right) spectra [Matricardi 2017].....	100

## **Table of Tables**

Table 1: EUMETSAT hyperspectral sounders characteristics.....	14
Table 2: List of IRS L2 products.....	22
Table 3: Definition of the PWLR <sup>3</sup> quality indicators.....	37
Table 4: Definition of the cloudiness summary .....	46
Table 5: List of stability indices in IRS L2 products .....	62
Table 6: Values of c1, c2, c3 (dependent on the value of t).....	65
Table 7: Constants for formula eq( 103) .....	76

## 1 INTRODUCTION

This document describes the Algorithm Theoretical Baseline for the processing of level 2 products from the EUMETSAT Infra-Red Sounder (IRS) mission to be flown on the Meteosat Third Generation (MTG) sounder platforms.

### **1.1 Purpose**

This Algorithm Theoretical Baseline Document (ATBD) introduces the mission and its intended applications. It describes the algorithms to be implemented in order to process the IRS Level 1 measurements and retrieve the required geophysical parameters at the required accuracy at Day-1.

## 1.2 Scope

This document is the baseline guiding the specifications of the operational IRS Level 2 (L2) processor. It is also intended to inform Users interested in understanding the algorithms and science underlying the generation of the operational IRS L2 products.

### **1.3 Applicable Documents**

[HQL2] MTG L2 products dissemination baseline EUM/MTG/DOC/09/0026

## 1.4 Reference Documents

[IRSL1ATBD] MTG-IRS Level 1 Algorithm Theoretical Basis Document EUM/RSP/TEN/16/878765

Scientific and technical reference documents are provided in Appendix A.

## 1.5 Document Structure

This document is divided into five sections:

## Section 1 General introduction - this section

Section 2 Short presentation of the IRS mission: its context, heritage, specificities and targeted applications.

### Section 3 Description of the algorithms sequence in the Day-1 baseline for operations

Section 4 Description of the generic algorithms, which can be called from different steps of the L2 processing sequence

## Section 5 List and short description of assumptions and open issues

Seven annexes are attached to this document:

#### **Appendix A: list of technical and scientific references**

## Appendix B: list of acronyms

## Appendix C: draft products content tables

Appendix D: apodisation, relative information content and sounding performances of IRS compared to IASI

Appendix E: demonstration projects and assessments results

Appendix F: discussion on radiance bias tuning

Appendix G: example of an implementation of slant to vertical coordinates computation

## 2 AN INTRODUCTION TO THE MTG-IRS MISSION

### 2.1 The Meteosat Third Generation Programme

In order to build on the success of the Meteosat First and Second Generation missions, EUMETSAT is developing the Meteosat Third Generation (MTG) satellites. After an elaborated user-consultation, the following needs have been identified:

- Continuation of the current imagery missions:
  - Full Disk High Spectral Imagery (FDHSI);
  - Rapid Scan Service (RSS).
- Development of new services:
  - Lightning Imagery;
  - Infrared Sounding.

To cover these needs, the MTG space segment will consist in six satellites of two different types, namely four imaging satellites (MTG-I) and two sounding satellites (MTG-S). The MTG-I hosts the Flexible Combined Imager (FCI) and the Lightning Imager (LI) instruments, while the MTG-S hosts the Infra-Red Sounder (IRS) and the Copernicus Ultra-Violet and Near-infrared sounder (UVN) instruments. In addition to the scientific missions, MTG will carry a small communications payload (GEOSAR, Geostationary Search and Rescue) to relay distress signals to a central reception station in Europe, for quick organisation of rescue activities.

### 2.2 MTG-IRS target applications

The IRS mission is primarily designed to support numerical weather predictions at regional and global scales, including nowcasting [MTG PP]. As concerns geophysical parameter products (aka Level 2 products), this is achieved through the provision in particular of:

- Temperature and moisture vertical profiles;
- Atmospheric Motion Vectors (AMV) derived from the former profiles;
- Instability parameters.

The instrument was hence specified with high spectral resolution in the infrared and high spatio-temporal sampling in order to provide frequent vertical atmospheric information to convective scale models, more consistently with their horizontal resolution. The high temporal frequency achieved from the geostationary orbit will in general increase the amount of information over dynamically important regions for Europe such as the North Atlantic and enhance the mid to short-range forecast capabilities.

The thermodynamic parameters and cloud information retrieved from the IRS observations have direct applications for nowcasting in complement to regional models outputs, with the aim to improve reliability and lead-time in identifying areas of interest, e.g. with rapidly developing atmospheric instability responsible for vertical motion, convection, precipitation and severe storms. The assimilation of hyperspectral sounder data in numerical weather models is based on level 1 radiances. Exploratory studies have started to evaluate the feasibility and advantages of assimilating alternative representation of Level 2 products in convective scale models. The temperature, humidity and ozone profiles can be derived in a much frequent and spatially resolved manner than with sounding missions on Polar orbiters, and hence be used as air-mass tracers to extract frequent 3D atmospheric motion vectors. These are in turn important inputs to both the regional and global numerical weather prediction models.

The secondary objective of the IRS mission is to support atmospheric composition and air quality monitoring, together with the UVN sounding mission. The IRS instrument is designed to acquire information about O<sub>3</sub> and CO. Other atmospheric compounds such as SO<sub>2</sub> and NH<sub>3</sub> are also expected to be monitored with MTG-IRS. Importantly, ozone may also be a tracer for atmospheric features and ozone profiles hence need to be generated for AMVs too.

### 2.3 The satellite platform

Contrary to the satellites of the first and second Meteosat generations which were spin stabilised, MTG-S and MTG-I will be three-axis stabilised platforms. The development of the MTG space segment is performed by Thales Alenia Space under ESA contract.

The satellite orbit is defined to be geostationary with a nominal altitude of 35786 km, an orbital period of 86164 seconds, an inclination of 0° (+/- 1°) and a Sub-Satellite Point (SSP) at 0°N 0°E. IRS: a hyperspectral sounder in GEO, measurements and acquisition principles

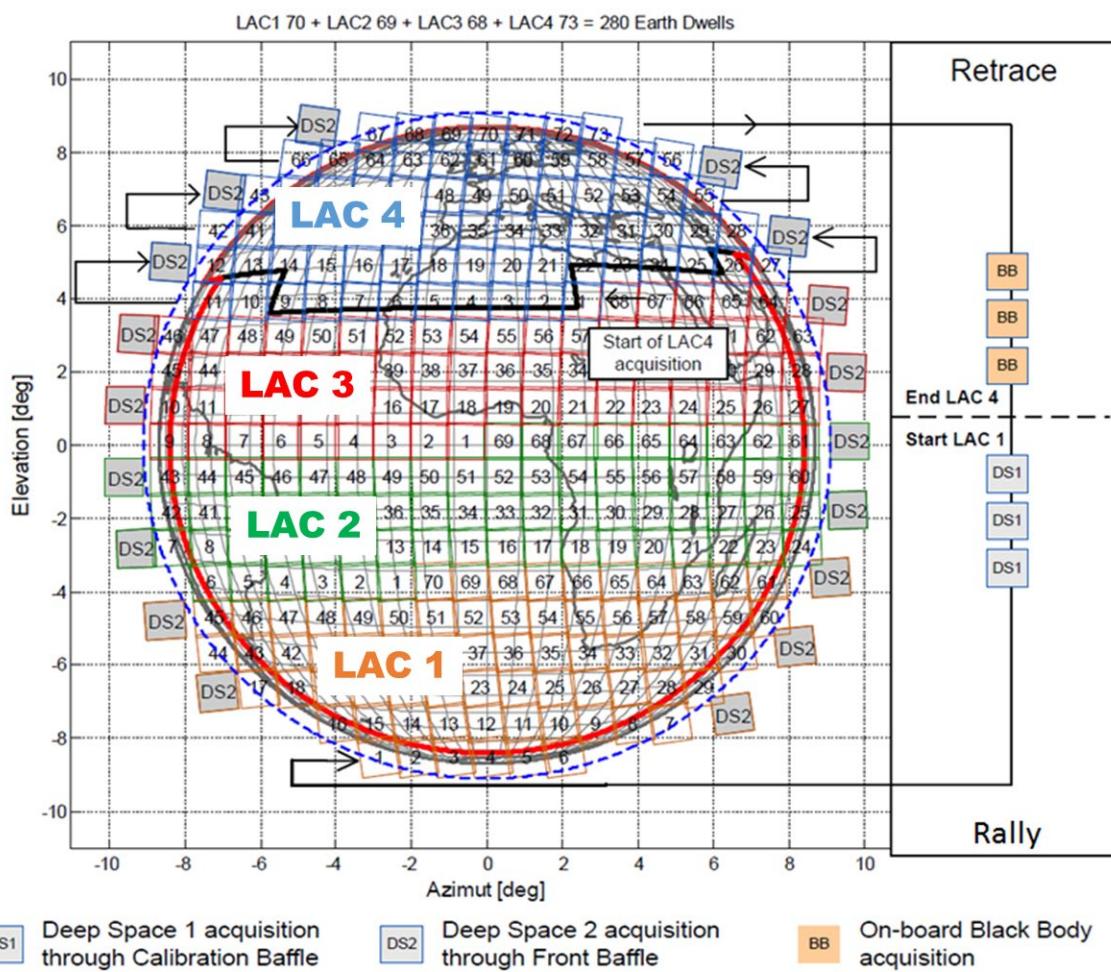
The IRS instrument is an imaging infrared Fourier-interferometer. It acquires a number of interferograms simultaneously over a so-called “dwell” using a two-dimensional detector array in two spectral bands: the LWIR (8.26–14.70 μm) and the MWIR (4.44–6.25 μm). The spectral soundings are transmitted to the ground as interferograms and transformed to radiances in spectral channels as part of the ground processing, before dissemination to the end users in calibrated Level 1 (L1) datasets. The reader is referred to the MTG-IRS L1 ATBD [IRSL1ATBD] for further details on the instrument itself and on the L1 processing algorithms.

As a result of this processing, 160x160 calibrated spectra are generated per band and per dwell, corresponding to simultaneous and contiguous observations within a dwell, with a spectral sampling of the order of 0.6 cm<sup>-1</sup>.

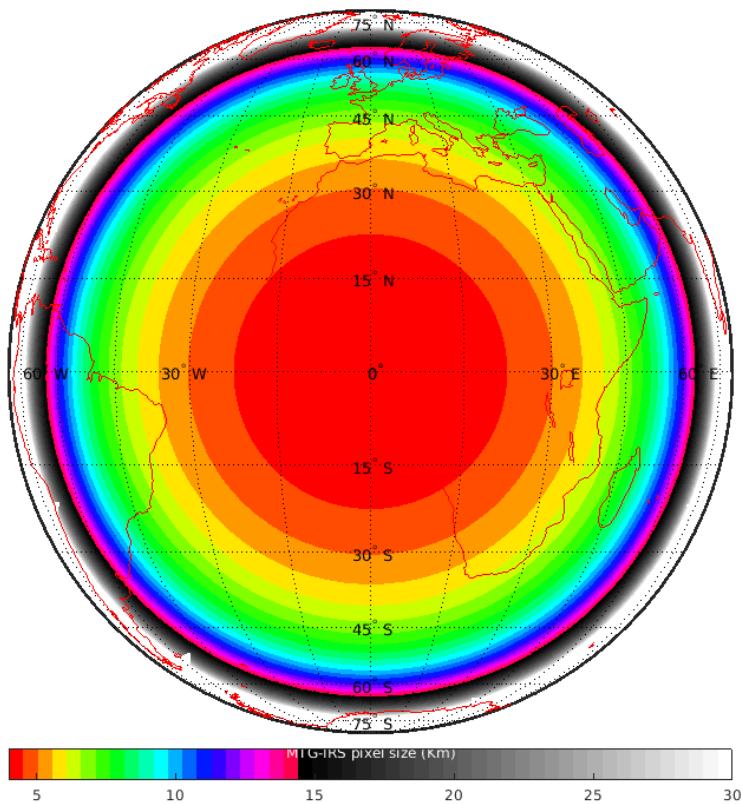
The Earth disk has been divided into 4 regions of interest called LACs (Local Area Coverage) numbered 1 to 4 from South to North, see Figure 1. IRS uses a ‘step-and-stare’ mechanism to observe the Earth in a contiguous manner. A single dwell is acquired within 9.7 seconds and is stepped in an East/West or West/East direction to form a line of dwell spectral soundings, before moving northward to form the next line. The operation is repeated until a LAC is completely covered, which will take 15 minutes.

The LAC 4, the northernmost LAC, is covering Europe and will be scanned every 30 minutes. The LACs 1, 2 and 3 will be imaged in-between successively. The nominal scan pattern, illustrated on Figure 1, has been defined to optimize the coverage and the need for calibration views. The scan pattern is configurable.

The field of view of one dwell is 1.025° and the spatial sampling on the Earth surface is 4 km at Nadir. The spatial sampling increases up to typically 7-8 km in Central Europe, 12 km in the Baltic region and exceeds 20 km as the instruments observes in the outer part of the disc, e.g. at higher latitudes in Europe(Figure 2).



*Figure 1: MTG-IRS dwells and Local Areas Coverage*



*Figure 2: Quickview of IRS pixel on-ground dimension [km]*

## 2.4 A major innovation with strong applicable hyperspectral infrared heritage

MTG-IRS will provide information on atmospheric thermodynamic parameters at an unprecedented vertical resolution combined with high horizontal (4 km at Nadir) and temporal sampling (30 minutes over Europe) due to its position in a geostationary orbit and hyperspectral information content. This represents a major innovation in an operational observing system and will be an important direct contribution to regional short-range weather forecasting and nowcasting, in complement to the traditional high spatial and temporal broadband imagery from geostationary orbit, which cannot provide as much atmospheric information in the vertical.

While the very first experience to date of a high resolution imaging-spectrometer in a geostationary (GEO) orbit is just being made with the FY-4/GIIRS Chinese mission, there is a long and strong operational heritage of hyperspectral sounding from space in Low Earth Orbits (LEO): e.g. with the EUMETSAT Polar mission IASI [Blumstein et al. 2004] or with the NASA/NOAA missions AIRS [Aumann et al. 2003] and CrIS [Han et al. 2013]. Their respective level 1 products belong to the same family of observations, i.e. highly spectrally resolved top-of-atmosphere calibrated infrared spectra. Assuming that the IRS L1 processing includes inter-pixel harmonisation and reduces instrument signatures, assuming also that an apodisation function can be coordinated between RTM developers and products Users, the IRS spectra from any pixel can be treated the same and a large part of the experience made in operational data assimilation and level 2 processing with AIRS, IASI and CrIS is directly applicable to IRS.

The present ATBD describes the IRS L2 Day-1 operational baseline algorithms, which expand upon the operational heritage of IASI L2 products processing at EUMETSAT. The rationale and motivation for this approach are manifold: to take advantage of the demonstrated operational experience with processing data of the same class, to ensure products consistency from EUMETSAT hyperspectral missions to the Users and to optimise scientific algorithms and software maintainability. There are also differences between these missions and specific aspects of MTG-IRS, which need dedicated attention to exploit the mission to its full extent in view of the targeted applications.

## 2.5 Specificities: limitations and opportunities

The IRS and its LEO predecessors have different acquisition principles, with viewing geometry spanning different ranges (reaching quasi limb in the case of IRS), yielding different spatial resolution and temporal sampling. Also, the observations are processed at different spectral resolution and coverage. The consequences in terms of opportunities, limitations or relative sounding performances expected for the IRS products services and the need to adapt heritage or study new algorithms are summarised in the following subsections. They are taken into account in the description of the present IRS L2 baseline for Day-1 and in the developments and study plan for future evolutions after instrument commissioning. The table below provides an overview of the main characteristics of the EUMETSAT LEO and MTG-IRS hyperspectral missions.

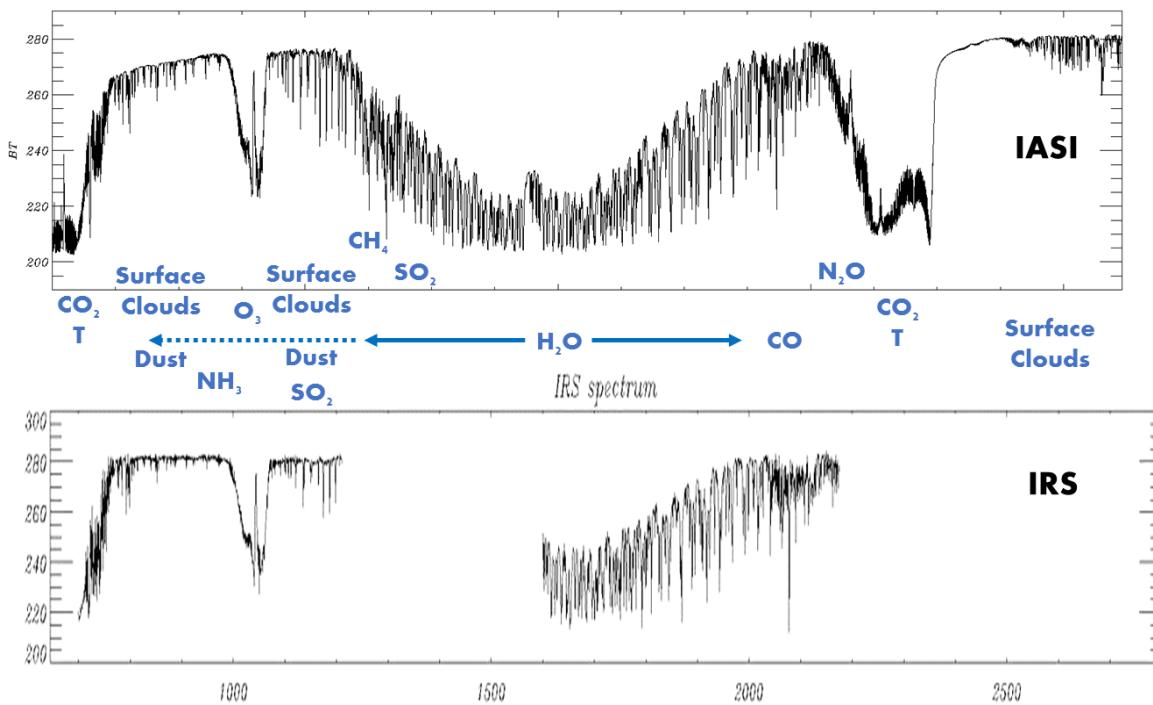
	<b>MTG-IRS</b>	<b>IASI</b>	<b>IASI-NG</b>
Orbit	Geostationnary Sub-satellite: 0°N;0°E	Low Earth Orbit Sun-synchronous 9:30 MLST	Low Earth Orbit Sun-synchronous 9:30 MLST
Instrument type	Michelson	Michelson	Mertz interferometer
MOPD	~0.828 cm	2 cm	~4 cm
Spectral sampling	~0.6 cm <sup>-1</sup>	0.25 cm <sup>-1</sup>	0.125 cm <sup>-1</sup>
Spectral coverage	LWIR: 700-1210 cm <sup>-1</sup> MWIR: 1600-2175 cm <sup>-1</sup>	645-2760 cm <sup>-1</sup> in 3 continuous bands	645-2760 cm <sup>-1</sup> in 4 continuous bands
Detector array	160 x 160	2 x 2	4 x 4
Pixel size	4 km (~7 km central Europe)	12 km (40 km swath edge)	12 km (40 km swath edge)
Spatial sampling	Contiguous pixels	Pixel separation ~20 km at Nadir	Pixel separation ~24 km at Nadir
Scan geometry	Overlapping images, from step and stare dwells From Nadir to quasi-limb views on disk edges (local zenith angle ~90°)	30 steps across track. From Nadir to ~60° local satellite Zenith angle	14 steps across track. From Nadir to ~60° local satellite Zenith angle
Acquisition frequency	Every 30 min over Europe. Full disk possible within 1h. 6h repeat cycle for full disk including the Southernmost LAC	Twice per day (more frequent revisit over the Poles)	Twice per day (more frequent revisit over the Poles)

*Table 1: EUMETSAT hyperspectral sounders characteristics*

### 2.5.1 Spectral coverage and resolution

Figure 3 shows the comparative spectral coverage of EPS/IASI and MTG-IRS together with the main geophysical information content. Because of the coarser spectral resolution and sparser coverage (see Table 1), the absolute sounding performances in terms of precision and vertical resolution will be slightly less with IRS than with IASI. A first theoretical assessment of relative performances for temperature, humidity and ozone is documented in . IRS can be expected to have one degree of freedom less than IASI in the troposphere for temperature and for water-vapour. Above 100 hPa, IRS will have only half the information content of IASI for temperature.

In addition to the thermodynamic profiles, the spectral coverage of MTG-IRS includes information about a number of atmospheric species. It is expected that ozone profiles from IRS are of slightly less but similar vertical resolution, nominally around 3 independent pieces of information and some relatively weak sensitivity in the lower troposphere. The IRS observations should allow also some quantitative estimates of the CO and NH<sub>3</sub> columns. Qualitative detection of SO<sub>2</sub> and ash in volcanic plumes is expected with IRS observations, using the weak  $v_1$  band of SO<sub>2</sub> (1100-1200 cm<sup>-1</sup>). However, quantification and height assignment for SO<sub>2</sub> and vertical resolution of CO profiles will only remain possible with IASI due to its larger spectral coverage (including the stronger SO<sub>2</sub>  $v_3$  band 1300-1400 cm<sup>-1</sup> [Clarisse *et al.* 2012]) and higher spectral resolution.



**Figure 3: Comparative spectral coverage of IRS and IASI**

### 2.5.2 No microwave instrument companion, higher sensitivity to clouds

The hyperspectral sounding Polar missions have been flying together with microwave (MW) sounders, (e.g. IASI with AMSU/MHS, AIRS/AMSU, CrIS/ATMS). These provide important

information in the presence of clouds as it is possible to sound the atmosphere within and below clouds in the microwave domain. Unlike its low-orbiting predecessors where synergistic use of MW and IR hyperspectral observations has been extensively made from LEO to maximise the yield and quality of the atmospheric soundings [IASI L2 PGS][NuCAPS ATBD][Susskind s. 2003], the MTG-IRS L2 processing will be IR-only, as MTG do not carry microwave sounding instruments.

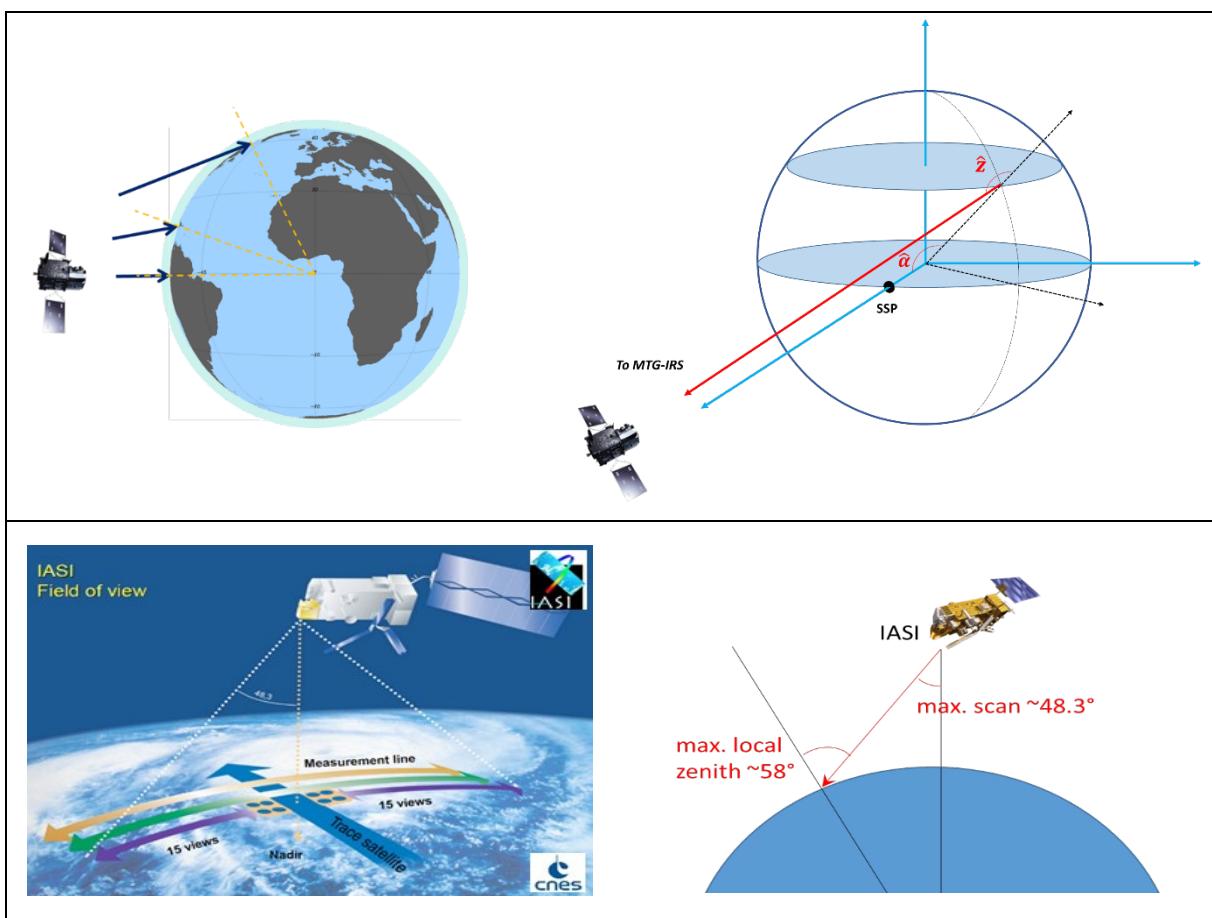
In the case of opaque clouds, only the cloud top and the atmosphere above account for the top of atmosphere radiances in the thermal infrared. Also the operational exploitation of pure infrared hyperspectral data in partly cloudy scenes or with semi-transparent clouds is still an active area of research and development because of the difficulty to model the cloud microphysics and radiative transfer therein. However, relevant experience has been made with the IASI L2 operational processor, which also has an IR-only fall-back mode when MW data are unavailable. Specific developments were carried out as part of IRS preparatory activities to consolidate and characterise the retrievals in the IR-only mode, which are summarised in EUM/RSP/VWG/17/948734. As expected, the yield with highest quality products is less in IR-only than in MW+IR, but IR-only retrievals of good quality are also possible in cloudy pixels with the IASI L2 operational algorithm.

### **2.5.3 Viewing geometry, from Nadir to quasi-limb viewing**

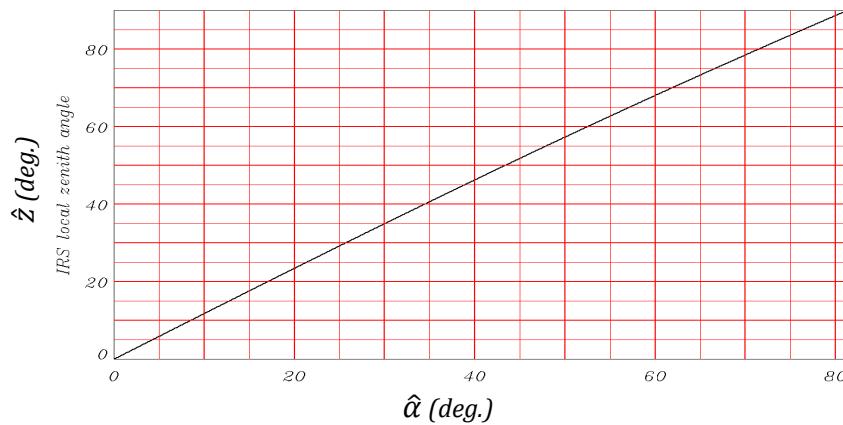
As illustrated in the Figure 4, EPS/IASI and MTG-IRS have a different viewing geometry. The local satellite zenith angle of IASI (and similarly for AIRS and CrIS) of targets on Earth ranges between Nadir and approximately 58° because of its swath geometry. MTG-IRS will however observe the Earth from a Nadir view at sub-satellite point (SSP) to a quasi-limb mode as the target point gets close to the limit of the visible disk. Figure 5 presents the variation of the IRS local satellite zenith angle ( $\hat{z}$  in Figure 4) as a function of the angular distance between the target point on Earth and the sub-satellite point ( $\hat{\alpha}$  in Figure 4). When IRS points along the Greenwich meridian,  $\hat{\alpha}$  is approximately equal to the latitude of the observed point.

The blue circle in Figure 6 shows the area where the local satellite zenith angles of IRS will be within the range of IASI's zenith angles or in other words, the range of viewing angles where the retrieval algorithms have been demonstrated in operations with the Polar mission. The orange and red circles indicate the respective limits of assimilation of MSG/SEVIRI radiances in numerical weather prediction models (up to 68° zenith angle) and of the production of operational global instability index (GII) products at EUMETSAT (up to 74° zenith angle). It is assumed that the IASI L2 processing concept can be applied ‘as is’ beyond the 58° local zenith but recent studies [Bormann 2016][Goukenleuque et al. 2017] have shown that the slant geometry should be taken into account when the local zenith angles reaches 60 to 65°.

In general, exploiting IRS observations at high slanted views –corresponding to a large portion of Europe- will require dedicated studies, for instance to account for possible variation of surface emissivity or to characterise the actual sounding capabilities (precision, vertical sensitivity and resolution) as Jacobians shift upwards and change shapes at high angles (Figure 7). The fast radiative transfer model RTTOV already includes spherical geometry and can be trained up to 85° local zenith angle. IRS-specific studies have been made already with a preliminary set of coefficients for IRS, including high zenith angles, as supplied by the NWP SAF on special request.

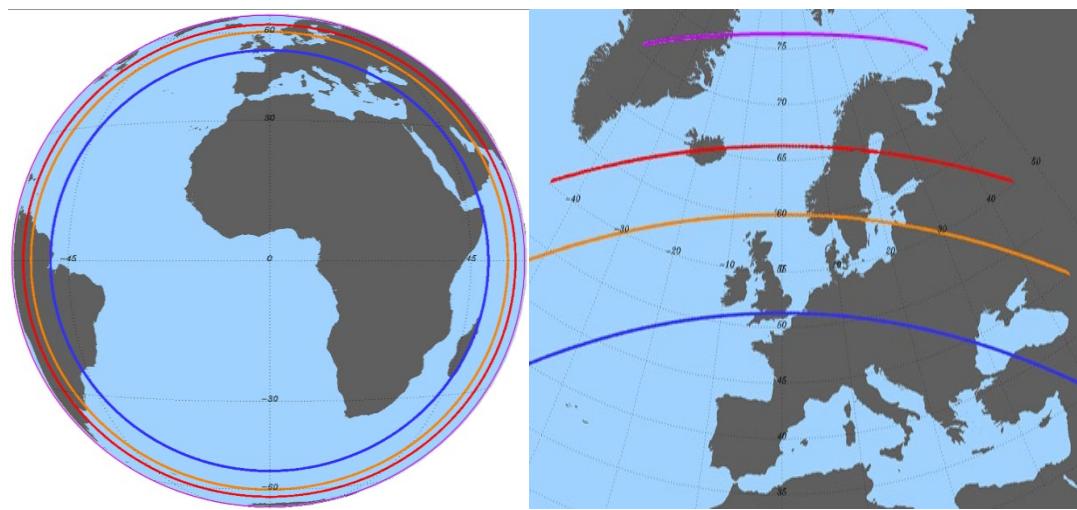


*Figure 4: Viewing geometry principles for IRS (top) and IASI (bottom)*

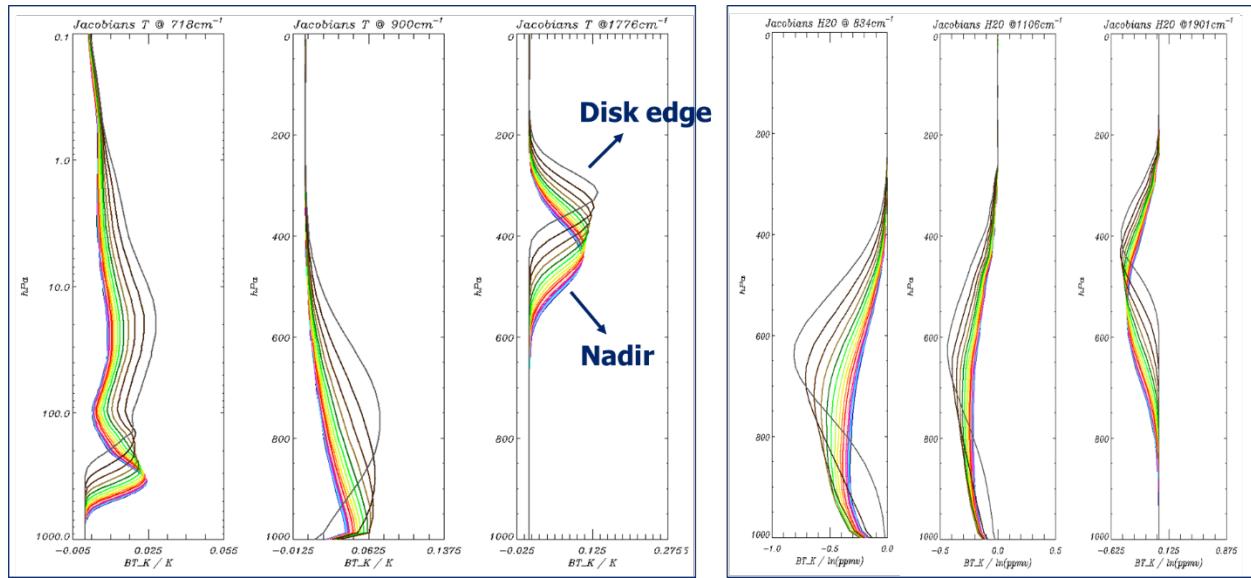


*Figure 5: IRS local zenith angle at target on Earth ( $\hat{z}$  in Figure 4) as a function of angular distance from sub-satellite point ( $\hat{\alpha}$  in Figure 4).*

Another complication with the IRS viewing geometry, as with any geostationary mission, is the parallax effect, which will not be negligible for a large part of the European domain. Although this is still ignored at this stage in the treatment of IASI at large scan angle and in the exploitation of SEVIRI in NWP or to generate SEVIRI GII or cloud products, a parallax correction is foreseen in the IRS L2 processing to provide atmospheric profiles at the vertical of the target point to the users (see section 3.10 and Figures 27 and 28).



**Figure 6:** Areas corresponding to the maximum local zenith angle with IASI (within the blue disk), used for assimilating MSG/SEVIRI radiances in NWP models (orange) and deriving instability indices from SEVIRI (red). The purple line shows the  $85^\circ$  zenith angle limit of applicability of RTTOV (nearly on disk edge in the left panel).

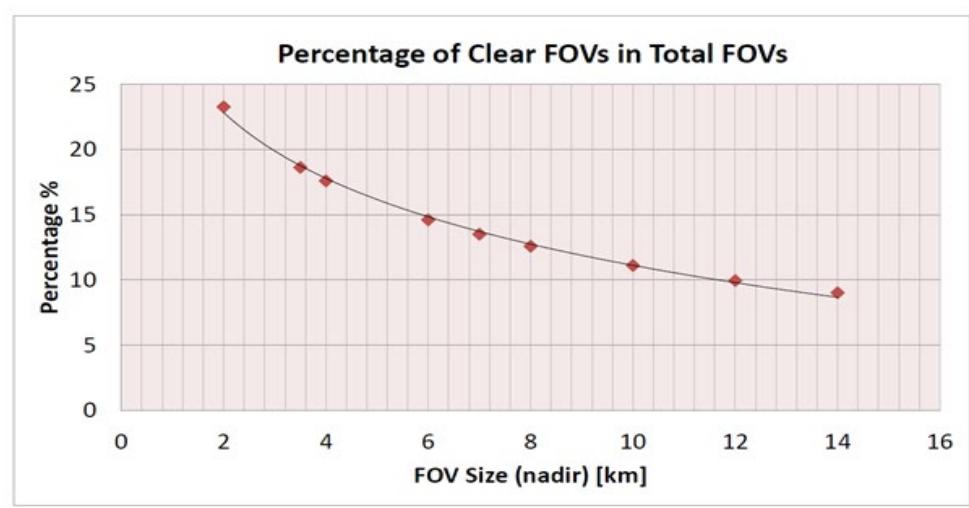


**Figure 7:** Examples of IRS Jacobians for temperature (left panel) and water-vapour (right) at different viewing angles.

#### 2.5.4 Spatial resolution and coverage

IRS will provide contiguous pixels (i.e. effectively forming images) covering entire scenes at high spatial resolution (4 km at Nadir) while the Polar hyperspectral missions, e.g. IASI or CrIS, have a coarser spatial resolution (12 km at Nadir, up to 40 km at swath edges) with also a sparse coverage (see illustration in Figure 9). The PWLR<sup>3</sup> algorithm (see PWLR<sup>3</sup>, §3.6.1.4), which exploits adjacent observations to make or regularise retrievals in individual pixels and was demonstrated in the IASI L2 processing is expected to be even more beneficial in the context of MTG-IRS. Furthermore, the data volume is higher by two orders of magnitude with IRS and requires CPU-effective processing algorithms, such as the PWLR<sup>3</sup>.

The variation of IRS pixel size within the visible disk is shown in the Figure 2, where it can be seen that the spatial resolution of 4 km at satellite sub-point in Africa will vary between 6 and 8 km in Western and Central Europe and reach about 12 km in the Baltic region. This and the spatial density with 160x160 adjacent spectra however represent major innovations as compared to the predecessor hyperspectral missions, whose footprint ranges in size between 12 km at Nadir to 40 km at swath edge.



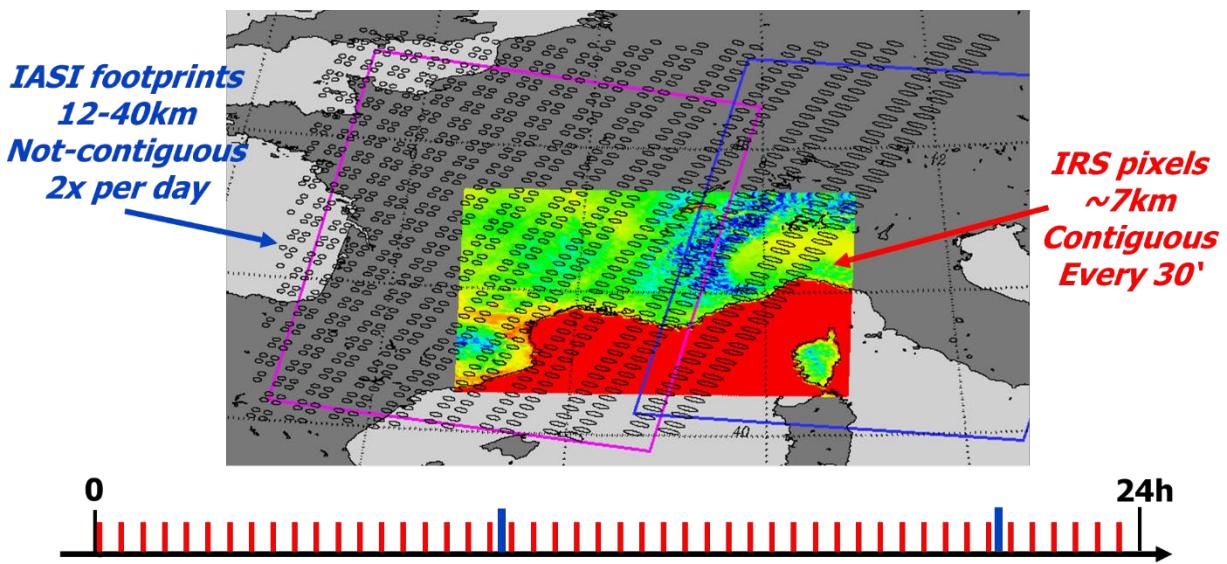
*Figure 8: Average cloud-free pixel percentage as a function of pixel size, from [Wang et al. 2016]*

This unprecedented spatial sampling from space with a hyperspectral instrument will be directly beneficial to resolve finer geophysical structures. In addition, compared to the LEO missions, smaller pixels such as IRS’ –and higher coverage density- increase the likelihood of sensing cloud-free scenes by 50 to 70% (see Figure 8, [Wang et al. 2016]), where hyperspectral IR L2 products are usually of best quality and where maximum information is assimilated in numerical models. The frequent sensing allows the look at atmospheric processes, which is a decisive assets of IRS and the main benefits of the mission.

### 2.5.5 Temporal sampling

As per User requirements, Europe (LAC 4) will be observed every 30 minutes, hence providing considerably more observations than from the LEO platforms, which revisit the same place only twice a day. Figure 9 sketches the spatio-temporal resolution of MTG-IRS and IASI acquisitions.

The frequent sensing over Europe increases the chances of clear-sky observations of a given point throughout the day. In addition, although not foreseen in the Day-1 baseline, it may be beneficial to exploit the frequent temporal sampling, for instance by using the retrievals from the previous scan as a starting point or *a priori* for the retrieval in the following. More sophisticated algorithms implementing Kalman filters could also be advantageously considered, following recent promising results with land surface emissivity retrievals from hyperspectral sounder data [Gray and Pavelin 2017]. Exploitation of the temporal information is considered a Day-2 activity and should include also possible developments of the statistical method in the temporal domain. It is noted that the frequent temporal sampling over Europe will also allow monitoring the self-consistency of the products with the previous observations.



**Figure 9: Simulation of an IRS scene using high resolution model data from Meteo-France, with realistic dwell limits (pink and blue) and IASI footprints from real observations in that area (black ellipses).**

### **3 DAY-1 PROCESSING SEQUENCE**

#### **3.1 Common HSIR L2 approach**

This MTG-IRS L2 ATBD builds on a common hyperspectral infrared (HSIR) Level 2 approach, which has been operationally demonstrated for IASI and is also the basis for the IASI-NG L2 processing. The rationale is to operate consistent L2 products from the three EUMETSAT hyperspectral sounding missions where commonalities exist and hence optimise use of resources during development, operation and maintenance of their processing algorithms and software. Despite differences between the three sensors (IASI, IASI-NG, MTG-IRS), a major part of the processing can be accomplished with generic software libraries applicable to all HSIR missions. Synergy can also be achieved in other technical elements for instance related to the off-line calibration of the retrieval algorithms, such as training software and databases.

In this general algorithm approach, a first retrieval is performed with a piecewise linear regression (PWLR) method at individual pixel resolution, under all cloud conditions ; Besides the geophysical state parameters, the function returns reliable quality indicators (uncertainty estimates). In the case of MTG-IRS, the method is intended to apply to all LACs for near-real time processing. This step is followed by a cloud detection and characterisation in terms of effective fractional coverage and cloud top height, with a variational method. A dust indicator completes the scene characterisation. In clear-sky pixels, a second retrieval is performed with optimal estimation method (OEM) with the objective to refine the retrieved quantities from the first retrieval.

In general, the MTG-IRS and Polar sounder level 2 algorithms have in common extensive application of principal components analysis (PCA) in the radiance and state vector spaces and the use of reconstructed radiances in variational methods to fully exploit the measurements while keeping dimensionalities low.

The present ATBD includes specific algorithms and configurations tailored to IRS. These are required by the nature of the mission, like the correction for the geometry (see §3.10.1) and due to its specific objectives, like the computation of convective parameters (see §3.10.3) or the configurable choice of the *a priori* in the variational methods (see discussion in §3.8.2).

#### **3.2 Processor inputs**

The main input data to the IRS L2 processing are the IRS L1 radiances represented as principal components (PC) scores, which is the baseline for NRT dissemination. The reader is referred to the MTG-IRS L1 ATBD [IRSL1ATBD] where the principles of the principal components analyses are recalled and where details of the specific processing for IRS are provided. Furthermore a static auxiliary dataset is used to configure the processing and a digital elevation model is used to derive the land fraction and surface elevation (mean and standard deviation) within each pixel.

Forecast fields from ECWMF are also taken as input to the processing and are collocated with the IRS pixels. They can optionally be replacing the PWLR profiles as *a priori* for the optimal estimation. They may be used also to flag the situations where numerical weather forecasts and

IRS L2 products significantly depart from each other, which could indicate areas where special attention is also required.

### 3.3 Output end-products

The end-products are organised in topical groups as summarised in the following table. The retrieved geophysical parameters and the estimates of the posterior error covariance matrix (from the OEM) are stored in separate products files. Preliminary products content tables are indicatively provided in Appendix C. They will be detailed in Products Format Specifications documents.

Short description	Content	Purpose
Atmospheric Temperature and Water-Vapour profiles	‘All-sky’ first retrieval + quality indicators Clear-sky variational retrieval	NRT Dissemination Archive
Surface parameters	Skin surface temperature over ocean (SST) and continental surfaces (LST) Land surface emissivity in selected channels (of the order of 10), from which the entire spectrum can be reconstructed by application of PC transformations	NRT Dissemination Archive
Cloud products	Cloud detection Cloud top pressure and effective fractional coverage Cloud phase Dust indicator	NRT Dissemination Archive
Atmospheric ozone profiles	‘All-sky’ first retrieval + quality indicators Clear-sky variational retrieval	NRT Dissemination Archive
Convective parameters	Collection of instability indices and integrated quantities to support the assessment of potential convective initiation	NRT Dissemination Archive
Error covariance matrix	Is the theoretical posterior error estimate computed after Rodgers [Rodgers 2000] for the parameters retrieved with the OEM: temperature, humidity, ozone	Archive

*Table 2: List of IRS L2 products*

### 3.4 Processing sequence overview

The five main processing steps are:

- I. Pre-processing (PREP)
  - Collocation of elevation and land/sea mask
  - Collocation of NWP forecast (optional *a priori* information)
- II. First retrieval, statistical method (PWLR)
  - Fast all-sky statistical retrieval of T, W, O, emissivity,  $T_S$  and cloud signal.
  - Retrieval includes reliable quality indicators (error estimates) for each of the retrieved parameters.
- III. Cloud retrieval/detection (CLOUD)
  - 1D-var retrieval of effective cloud fraction and cloud top pressure of up to two cloud layers using the simple cloud parametrisation of RTTOV as forward operator and analytic expressions for optimal effective cloud fractions for given cloud top pressures.
  - Cloud phase by simple brightness temperature difference tests
  - Dust index by linear regression
- IV. Second retrieval, optimal estimation (OEM)
  - Clear sky optimal estimation of T, W, Ozone, (CO<sub>2</sub> as optional by-product of T), emissivity.
  - Measurement represented as channel subset of reconstructed radiances additionally filtered by forward model reconstruction to suppress instrument artefacts
  - PWLR retrieval or collocated NWP forecast used as *a priori* (and first guess)
- V. Post-processing
  - Reconstruction of the profiles at the vertical of the target point on Earth.

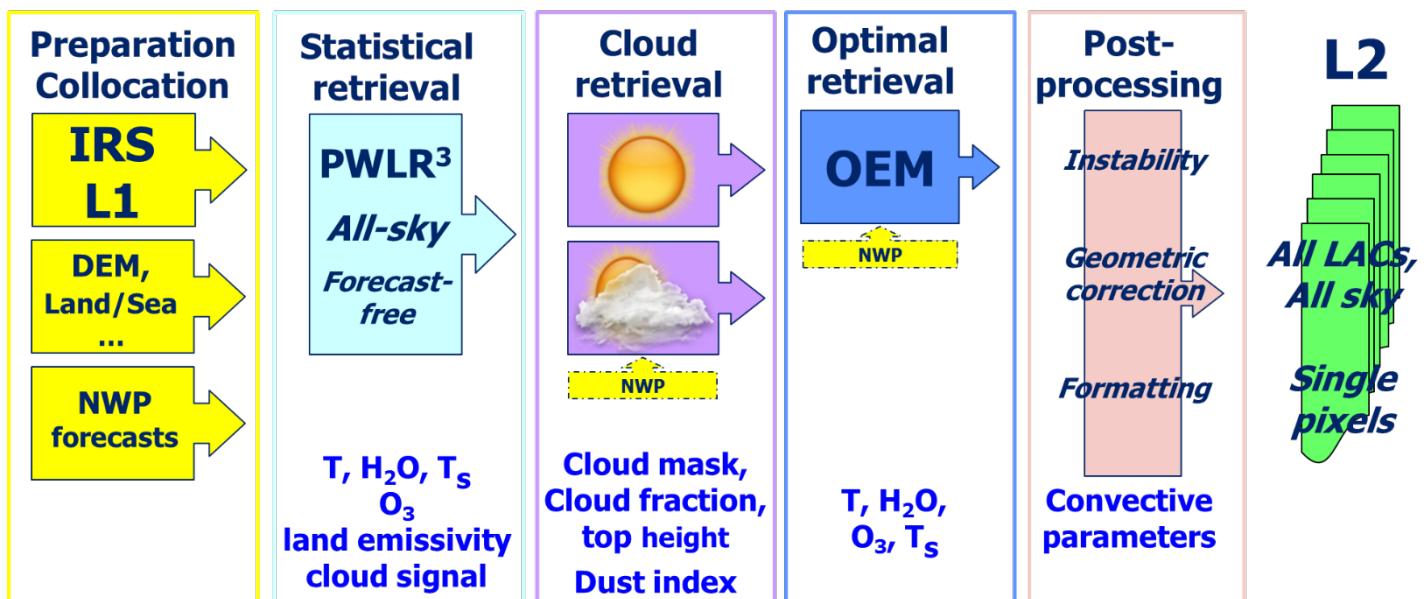


Figure 10: High level overview of IRS L2 processing steps – the optional use of NWP forecasts in the cloud and OEM retrieval steps instead of PWLR<sup>3</sup> profiles is indicated with the dash yellow box

## 3.5 Pre-processing

### 3.5.1 Input preparation and collocation

The purpose of the input data preparation and collocation is to gather MTG-IRS measurements and relevant collocated data in a common file, which serves as input for the further sub-functions.

The collocated data include ECMWF forecast data as well as land fraction and surface elevation mean and standard deviation within each MTG-IRS field of view.

Most of the data are organised according to IRS scan line (LINE) and pixel number (PIX). The collocated ECMWF forecast data are an exception, here the fields are kept at the original ECMWF grid points, but only grid points within a rectangular area covering the current IRS dwell are retained. This allows for posterior application of different method of collocation, such as nearest neighbour vs bilinear interpolation or along slant path or vertical. The collocation of ECMWF forecast fields is detailed in the section 3.5.4.

### 3.5.2 Acceptance of L1 measurements

This function identifies the pixels which can be used for retrieval by examining the quality flags of the Level 1 processing and checking that the measurements as well as the geolocation and sun and satellite angles are within the expected bounds.

The input to the IRS L2 processor are the IRS principal components scores (PCS) products, as disseminated in near-real time to the users. The L2 retrieval functions requiring principal components scores as inputs will directly take these scores from the IRS PCS products.

For the L2 retrieval functions requiring radiances in selected channels, a reconstruction of the IRS radiances from the PCS in the corresponding channels will be performed as explained in [IRSL1ATBD]. The reconstruction operator from scores to radiances shall include the linear transformation to compute apodised radiances consistently with the radiative transfer model used in the L2 processing. A reconstruction operator including same apodisation as in RTTOV-IRS is planned by NWP SAF in the IRS-PP, the IRS post-processing package [IRS MAG 5].

### 3.5.3 Collocation of Digital elevation model and land-mask

This step is performed in order to produce the mean and standard deviation of the surface elevation within each pixel as well as the land fraction. To avoid inconsistency the same atlas should be used for the land fraction and the surface elevation. To be in line with the specification for IASI-NG, we propose to use the ACE-2 30" database [ACE-2], which provide surface elevations on a global longitude/latitude regular grid with a resolution of 1/120 degrees and in which sea surface locations are distinguished with a special value (-500). However another atlas may be considered depending on the experience made with this atlas and on potential drivers for common atlases in the MTG programme.

As a simple approximation for this purpose, we will assume that the IRS point spread functions are rectangular, uniform, contiguous and non-overlapping. In this way the point spread functions (PSF) can be regularly sampled in a (LINE, PIXEL) coordinate system, where all points have equal weights. In each dwell the geolocations of all pixels are transformed into Cartesian coordinates and used to derive by regression a linear relation between the (line, pixel)

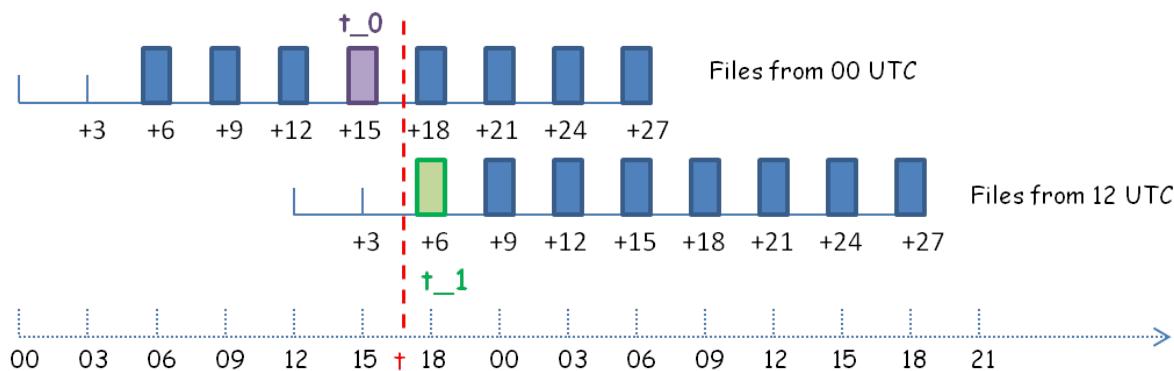
coordinates and the geolocation. This allows to get the longitude and latitude of all PSF sample points (after conversion back from Cartesian coordinates) such that they can be looked up in the ACE-2 atlas. Then it is a simple matter to compute the mean and standard deviation of the surface elevation as well as the fraction of the field of view which is covered by land.

### 3.5.4 Collocation of NWP forecasts

In this step the ECMWF NWP forecast data within a rectangular area fully covering the current dwell are extracted and written to the PREP file at the original horizontal and vertical sampling closest to IRS resolution. The relevant variables are:

1. Temperature
2. Specific humidity
3. Ozone Mass Mixing ratio
4. Skin temperature
5. 2m Temperature
6. 2m Dew point temperature
7. Surface pressure
8. 10 m U-velocity
9. 10 m V-velocity

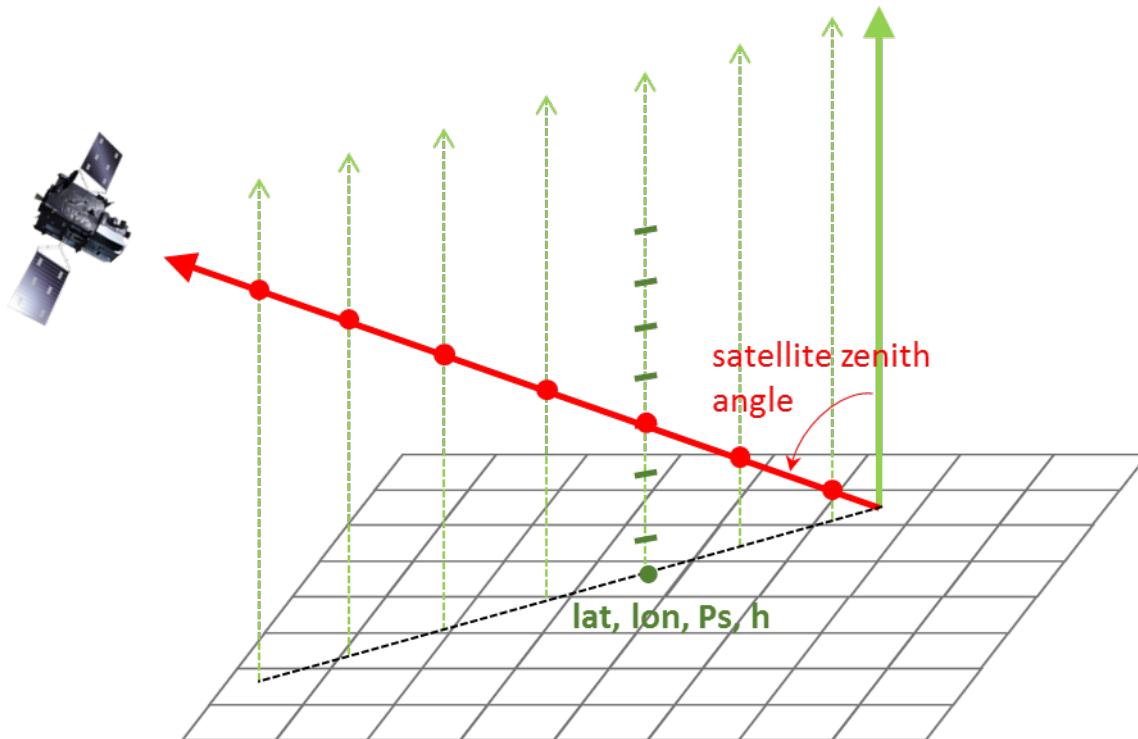
It is planned to make use of the best affordable forecast temporal and horizontal sampling, using EMCWF fields in their full vertical resolution on hybrid levels. Based on the experience with EPS, 3h-forecasts at 0.125° sampling in the horizontal are planned at the minimum, with ECMWF data files being delivered at step 06, 09, 12, 15, 18, 21, 24 and 27 twice a day (from the two analysis cycles at 00 and 12 UTC), as shown with boxes in Figure 11. In this figure, the red line corresponds to the sensing time ( $t$ ) in the timelines. The closest in time surrounding forecasts from the most recent analysis are linearly interpolated in time to the sensing time ( $t$ , in red in the figure). In this example, the first surrounding forecast ( $t_0$ , violet in the figure) was not available from the 12UTC analyses and is taken from the older analyses. The second surrounding forecasts ( $t_1$ , green in the figure) is available from the most recent analyses.



**Figure 11 ECMWF forecast files sequence and selection for interpolation to sensing time  $t$  (red line).**

The output variables written to the PREP files are obtained from two forecast fields by linear interpolation in time. These two forecast fields are chosen such that one has validity time before the sensing time of the granule and the other has validity time after, while both having validity time as close as possible to the sensing time. In case of draws, the field with the most recent base time is preferred.

The collocated forecast data are to be used online if the L2 processor is configured to use forecast as *a priori* in the optimal estimation. The profiles should be extracted along the slant paths (see Figure 12) from the data retained in the PREP files, to avoid errors of using a vertical profile in the calculated TOA radiances [Bormann 2016] and [Goukenleueque et al. 2017].



**Figure 12: Observed slanted profile (red) and vertical ID profile (green) as stored in the forecast data.**

The construction of the slanted profile is performed as follows:

- i. Determine the surface pressure and temperature corresponding to the centre of the IRS pixel, from the nearest forecast fields.
- ii. Determine the altitude of the pressure levels used in the OEM retrieval with RTTOV as explained in the pressure to height conversion in §4.3, with the inputs from the above point and the altitude from §3.5.3.
- iii. Determine the geolocation at the vertical of these pressure levels as explained in §4.16, with the altitudes computed above, the satellite viewing angles, the pixel geolocation and the surface elevation.
- iv. Identify the nearest grid point in the forecast grid for the sub-profile points and interpolate the vertical profiles from the nearest forecasts to the corresponding pressures, as described in §4.1.

In order to minimise computation costs and assuming that the spatial sampling of the forecast fields is at least of  $0.125^\circ$ , the horizontal collocation in the steps *i* to *iv* can be performed with the nearest neighbour.

### 3.6 First retrieval: PWLR<sup>3</sup>, an “all-sky” 3D statistical retrieval

This section describes the first retrieval, which is based on a statistical algorithm. It first introduces the concept and rationale behind the algorithms choices (3.6.1) and presents in

---

details the actual retrieval of the geophysical parameters and associated quality indicators as well as the method and dataset to compute the algorithm configuration (3.6.2).

### **3.6.1 Algorithm concept and rationale**

#### **3.6.1.1 PieceWise Linear Regression: a non-linear statistical retrieval**

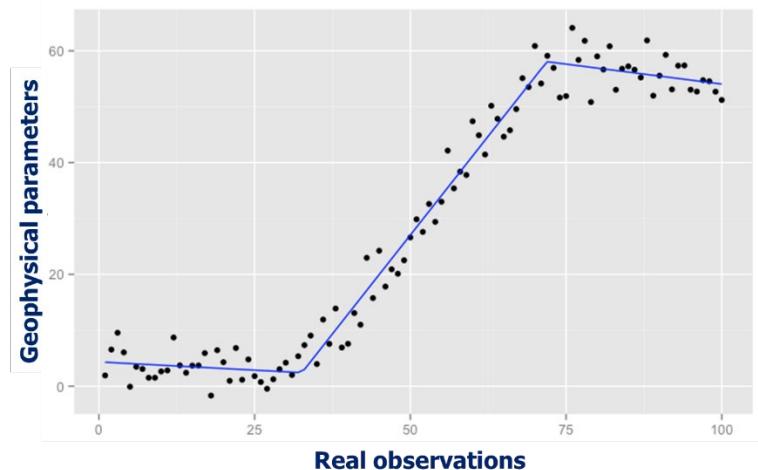
Statistical retrieval methods model the statistical relationship between the observations and the parameters to be retrieved based on a large representative dataset. They are popular inverse methods for atmospheric sounding, with long heritage [Smith and Woolf 1976]. Statistical methods are intrinsically much faster than so-called physical methods, where the physical processes in the radiative transfer are modelled in parametric equations and inverted directly or in an optimal way. The latter methods require explicit forward computations with a radiative transfer model (RTM) as part of the retrieval process, and RTMs are usually the bottleneck in terms of floating-point operations.

The top of the atmosphere radiances observed by hyperspectral sounders are a complex function of the geophysical parameter state vector to be retrieved. Furthermore, the problem is ill-posed. Different state vectors can result in the same TOA observation, there is hence not a unique state vector corresponding to a given observation. Often, a statistical method has been used to provide a first-guess state vector to an optimal estimation method [Zhou et al. 2007], [Susskind s. 2003], [August et al. 2012], [NuCAPS ATBD], solving the non-linear problem in an iterative manner [Rodgers 2000].

In some cases the relationship between the observations and the geophysical parameters is modelled with linear functions, possibly distinguishing different sensing regimes e.g. under clear/cloudy conditions or at different viewing angles [Zhou et al. 2005], [Smith et al. 2012]. The linear model is obtained in a so-called training phase, by regression between the predictors, usually consisting of the measurements represented as principal component scores, and the target geophysical parameters. Artificial neural networks (ANNs) are another kind of statistical methods, which have also been extensively used to retrieve geophysical parameters from hyperspectral sounders [Hadji-Lazaro et al. 1999], [Blackwell et al. 2005], [Crevoisier et al. 2009], [Han and Sohn 2013], [Susskind et al. 2014], [Whitburn et al. 2017]. Like the linear methods, ANNs are statistical regressions between inputs and outputs, with the additional possibility to model non-linearities in the relationship as observed in the training sample [Le Cun et al. 1998].

With the piecewise linear regression (PWLR), the overall non-linear relationship between observations and state vector parameters is approximated by multiple linear models, in different classes. If the classes are sufficiently small, this retrieval method can be shown to be equivalent to optimal estimation: “Thus linear MAP [maximum *a posteriori*] solution can be regarded as a simulation of a multiple regression method where coefficients are found which best relate the state to the measurements” [Rodgers 2000]. Figure 13 illustrates the concept of piecewise linear regression, specific detailed description of the clustering in this context is provided in section 3.6.2.5. The leading principal components scores of the measurements are used as input to the PWLR retrieval. This is to remove the collinearity, which would be present if all channel were to be used as predictors, and also consequently reduces the dimensionality (typically by one to two orders of magnitude). As compared to using a channel subset, this also reduces the influence of instrument random noise, which is filtered to a large extent by excluding the higher order principal components from the retrieval [IASI PCC val]. It was introduced in the release of

the EPS/IASI L2 operational processor version 6 in 2014 and has proven a very accurate, precise [Boylan et al. 2016], [IASI L2 v6 val], [Roman *et al.* 2016], [Feltz et al. 2017] and fast retrieval method.



**Figure 13:** Piecewise Linear Regression concept. Linear relationship is modelled in different classes of observations, here illustrated with the 3 segments. In real product processing, the clustering is based on the observations by application of k-mean techniques.

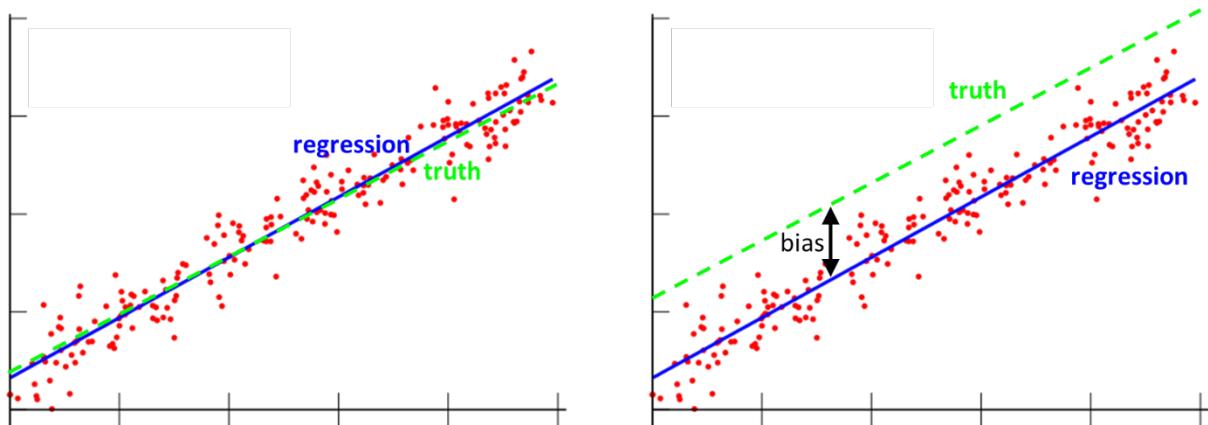
### 3.6.1.2 Training with real observations

All statistical methods require an offline training phase, where the statistical relationship between predictors and target parameters is inferred from a large set of correlative input/output pairs. The training set must be representative of the range of atmospheric situations that the instrument will sense during its operations. One approach consists in building a synthetic training set with a representative geophysical database, like for instance [Chevallier 2002], and in simulating the corresponding radiances with a radiative transfer model. Errors associated to the forward model become a limitation with this approach. These are for example coming from underlying spectroscopic uncertainties, inaccurate representation of the surface emissivity properties and in general the variation of static (i.e. non-retrieved) species which are not taken into account in the training simulations. Also, importantly, forward modelling of hyperspectral infrared cloudy radiances is still an active subject of research and the uncertainties in RTM cloudy simulations are usually much higher than in clear-sky.

Aiming at accurate retrievals in clear and cloudy scenes, the piecewise linear regression approach for hyperspectral sounders at EUMETSAT is trained with real observations associated with the best available co-located description of the atmosphere. Typically  $10^7$  to  $10^8$  teaching pairs are used in the training, from randomly selected days covering whole seasonal cycles. In the EPS IASI operational context, they consist for instance of IASI observations with collocated re-analysis fields from ECMWF. If the relationship in a regression class consisting of similar observations is close to linear and the collocated fields have random errors, it is natural to expect that the resulting regression will be more precise than the training target set, which is illustrated in the Figure 14. If however the training base was biased with respect to the truth, the statistical method trained with such a dataset would inherit also the same bias (see Figure 14, right-hand side). This can be avoided by correcting the training set for known biases, prior to the regression, as it must be done for the training of the land surface temperature (see details in 3.6.2). It must be noted that regression trained with synthetic data

may introduce biases into the retrievals, when biases between observed and calculated radiances (see Appendix D) are not properly accounted for.

For IASI, 200 PC scores in total are used for regression in different regression instances. These numbers are determined empirically at this stage.

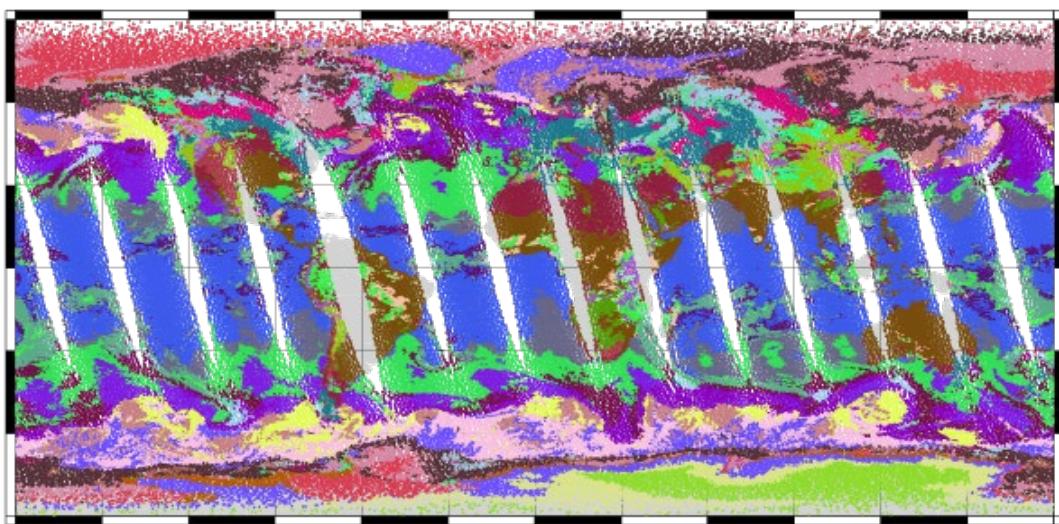


**Figure 14:** Linear regression (plain blue) modelling true (dashed green) function from experimental collection of input/target pairs (red dots) in an unbiased (left) and a biased (right) sample.

### 3.6.1.3 Defining the regression classes

Different sets of coefficients are often derived for different retrieval classes, for instance for different viewing angle classes [NuCAPS ATBD] or different day/night or regional classes [Milstein Blackwell 2016], [Jang et al. 2017]. While such stratifications intend to specialise the retrieval function to some observation regimes and hence to seek better performances, there is a risk of spatial discontinuity of the retrievals at the class edges, e.g. at latitude bands borders. Furthermore, a training set “specialised” for a given regional box may actually include very different atmospheric situations (e.g. dry hot scenes and cloudy atmospheres, high-shooting clouds...), which will limit the validity of a linear regression model but also the effectiveness of non-linear inference due to the large heterogeneity of the scenes.

Instead, the present PWLR approach relies on a fine stratification which is based on the observations themselves. This aims at grouping meteorological situations corresponding to similar types of observations in the same training subsets, regardless of their actual location on Earth or time of the year. Within these subsets of the overall training set, we can see that the linear model is a more valid approximation. The classification itself is obtained with the k-mean clustering technique [MacQueen] applied to a subset of the leading PC scores. Figure 15 illustrates the result of such a clustering with IASI data from ascending nodes, where each colour corresponds to a particular class identifier.



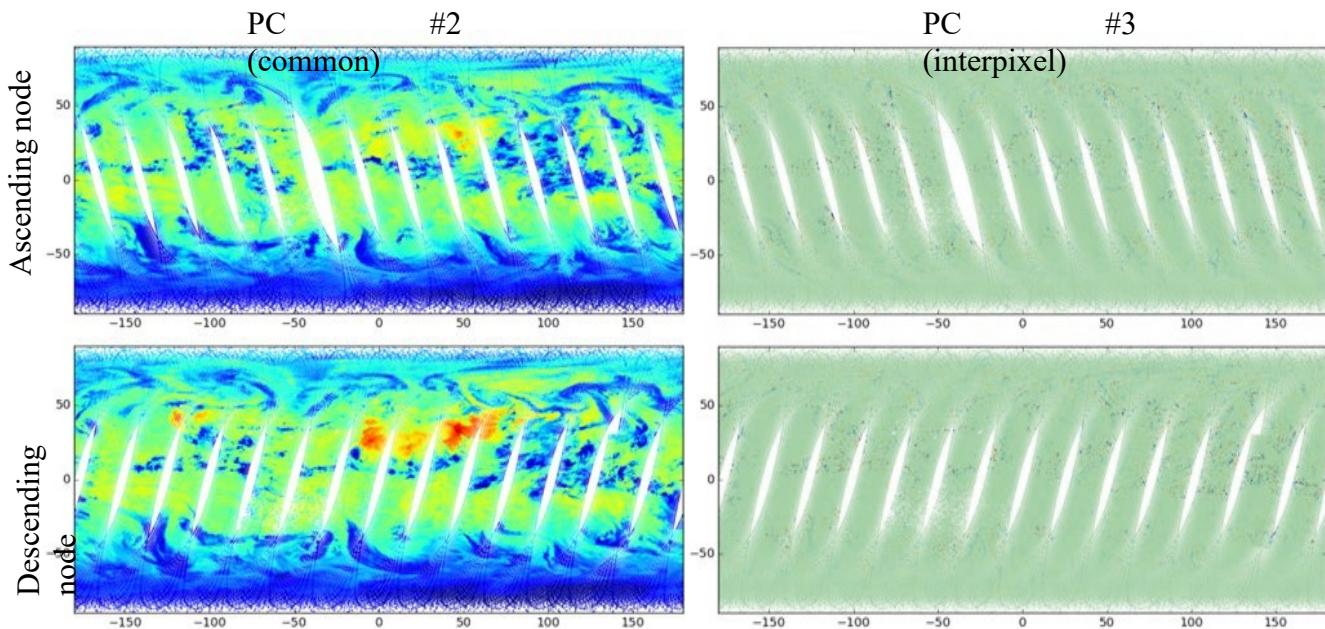
*Figure 15: An example of PWLR clustering based on IASI observations.*

To reduce random errors, the overall retrieval is obtained by averaging a number of ensemble retrievals, which are based on different instances (or realisations) of the coefficients. These instances are all similar, but clearly in order for the ensemble approach to be beneficial, they must be different and result in retrievals errors which are not fully correlated among the instances. This is achieved by varying the number of predictors (and possibly their relative scaling) which are used in the clustering and/or by using different subsets of the overall training set. For instance for IASI in IR-only model, four different regression instances are obtained by clustering using the first 6 and 9 PCs in the common directions to the adjacent pixels (see §3.6.1.4) and with two halves of the overall training set.

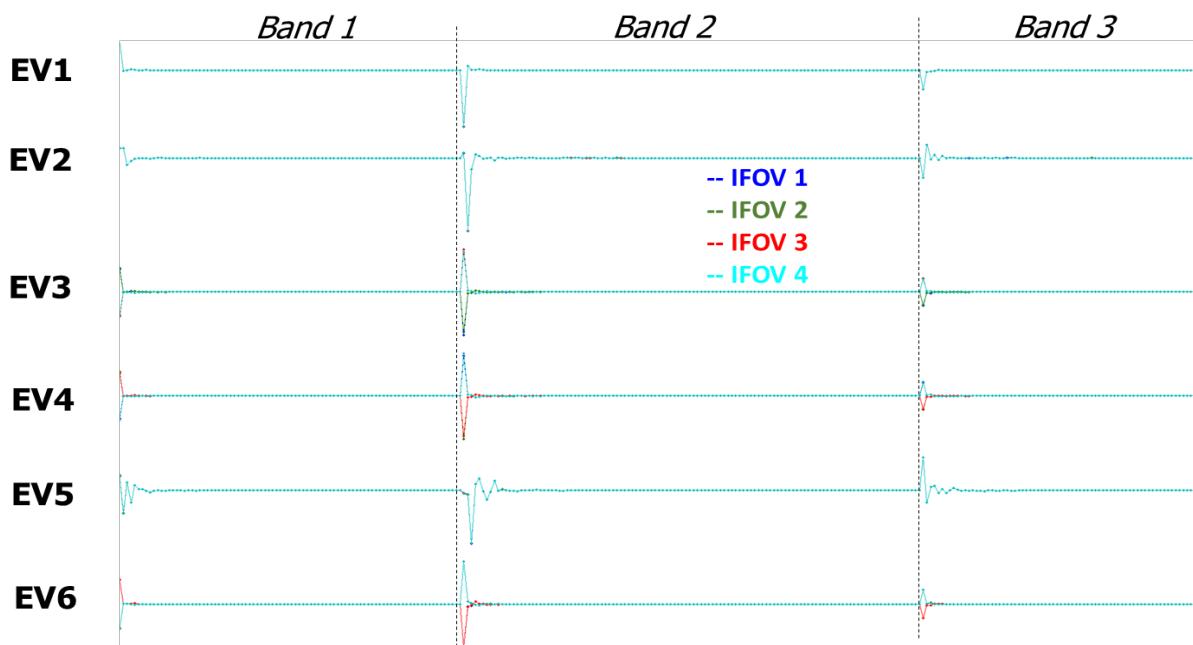
### 3.6.1.4 The PWLR<sup>3</sup>, a 3D retrieval

The PWLR<sup>3</sup> (PWLR-cube) is the latest evolution of the PWLR algorithm tailored for IASI [Hultberg and August 2015]. It exploits measurements from a group of adjacent pixels to perform retrievals in each of the individual pixel. The rationale is to take advantage of the horizontal correlations that exist in the atmosphere to bring additional information in a given pixel from its neighbourhood. Let us consider for instance the situation where some pixels are fully or partly covered by clouds. The information below the cloud level coming from the clear-sky or from less cloud-affected adjacent pixels can be exploited. The 3D correlations, which are exploited in this approach, have a regularising effect in the vertical as well as in the horizontal on the solution across all adjacent pixels.

It is interesting to note that the principal components corresponding to the measurements of a group of adjacent pixels can be clearly divided into two different types, which either capture features which are common for all the pixels or capture interpixel differences. In Figure 16 and Figure 17 the two type of PCs are illustrated.



**Figure 16:** Maps of principal component scores in common (in PC #2, left) and interpixel differences (in PC #3, right) directions of 4 adjacent IASI pixels.



**Figure 17:** Eigenvectors for IASI pixel (IFOV) quadruplets, carrying common (EV1-2) and interpixel (EV3-6) information.

This concept was introduced in the IASI L2 PPF v6.2 released in June 2016 and proved to yield more precise retrievals than the single-pixel-based PWLR approach [IASI L2 v6.2 val], [Sun *et al.* 2017]. The horizontal correlation of the atmospheric state vector being higher at smaller scales, it is expected that the benefits of exploiting adjacent information will be even more important for a spectro-imager like MTG-IRS, which has small and contiguous pixels as compared to IASI, whose pixels size and separation range between 12 and 40km.

The practicalities of applying the PWLR<sup>3</sup> concept to IRS will be further studied, for instance to determine the size of the window where adjacent measurements are exploited, to determine the treatment of non-nominal cases (e.g. dwell borders, dead or bad pixels...). It is also envisaged with this approach to retrieve directly the atmospheric profiles at the vertical of a pixel centre and hence possibly to avoid reconstructing this information from slant retrieved profiles as discussed in section 3.10.

### 3.6.2 Detailed description of the algorithm and configuration

#### 3.6.2.1 Training set

The training set is composed of real measurements paired with collocated reference data for the parameters to be retrieved, which include:

- temperature profiles (T),
- water-vapour profiles, specific humidity (W)
- ozone profiles (O),
- surface skin temperature ( $T_s$ )
- land surface emissivity E
- surface pressure ( $P_s$ )
- cloud signal (OmC, stands for Observed minus Calculated).

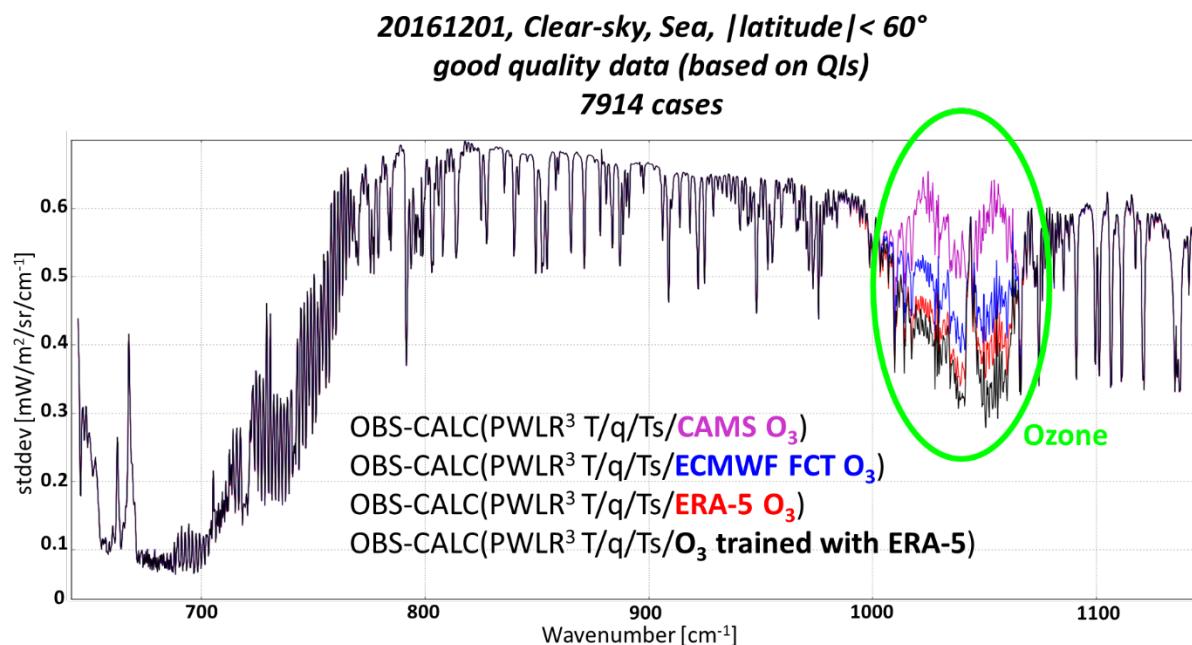
To avoid overtraining when using a large number of regression classes and to avoid sensitivity to random errors in the reference data it is important that the training set is proportionally larger, by several orders. While it is hard to be more precise on the number of pairs (i.e. predictor/geophysical parameter) required in the training set, it is possible to check that the generated coefficients do not suffer from overtraining by applying them to independent data. As an example, the training of the IASI L2 PPF version 6.4 was done with 96 full days of IASI measurements spread out over two years, 2015 and 2016, for a total of about 100 million of training pairs of measurements and targets.

The choice of the reference data to be paired with the measurements is important, since any systematic biases present in the reference data are inherited by the PWLR<sup>3</sup>. While the PWLR<sup>3</sup> method is sensitive to systematic errors in the reference data it is largely insensitive to random errors (see Figure 14), provided that the training set is big enough, e.g. statistically representative of the error probability density function. It is perfectly normal to achieve a retrieval precision which is better than the precision found in the training data.

For the latest IASI training, the reference data of T, W and  $P_s$  is taken from the ERA-5 reanalysis which incorporates the most recent model updates and is available at high temporal resolution. Currently the ERA-5 data are only available until the end of 2016, but it is foreseen that a continuous update providing so called ERA-5T data with a short delay of less than one week will be set up soon [ERA-5]. It would hence also be possible to use ERA-5 (or follow-up project) for training of MTG-IRS PWLR<sup>3</sup> once the IRS measurement become available. The T and W profiles are retrieved at the 137 model levels at which they are provided in the ERA-5 data with units of K and kg/kg respectively. W profiles are handled in log(W) for training and retrieval purposes.

The Combined ASTER and MODIS Emissivity over Land [CAMEL] database from University of Wisconsin is used as reference for emissivity. In the PWLR<sup>3</sup> regressions, the emissivity is represented and retrieved in a limited number of principal components (typically less than a dozen), as done for instance in [IASI L2 PGS], [Pavelin and Candy 2013], [Zhou et al. 2011]. The full land surface emissivity spectra can then be expanded from the retrieved principal component scores; in particular at the selected channels needed in the subsequent retrieval steps (e.g. clouds or OEM) or to be written in the final IRS L2 product. Over maritime surfaces, no surface emissivity will be generated as a product. The analytical expression of the ocean surface emissivity built-in RTTOV is intended instead in the subsequent retrieval stages.

The reference data for training the ozone (in log(kg/kg)) is also taken from ERA-5 for now, as ozone fields from the Copernicus Atmosphere Monitoring Service [CAMS] did not yield as good a radiance fit to real observations. This is illustrated in the Figure 18 showing the standard deviation of clear-sky IASI observations (OBS) minus calculated (CALC) radiances over ocean, computed with different ozone sources and the same PWLR<sup>3</sup> retrieved temperature and water-vapour profiles. The evolution of the two models for ozone will be monitored to guide the choice of the actual training set by the time of MTG-IRS launch.



**Figure 18:** IASI observations (OBS) minus calculated radiances, using different collocated ozone sources (purple: CAMS analysis, blue: ECMWF IFS forecasts, red: ERA-5, black: retrievals based on ERA-5)

This will be part of Commissioning activities to determine the first version of the coefficients for operational production. A rolling training mechanism is planned with updates for instance after typically a day, a week, a month, a season and a year. After a full seasonality, it is expected that only longer term updates will possibly be needed. This will be part of the product monitoring to continuously assess the applicability of the coefficients and trigger updates when needed. The lessons learnt from Metop-C, where a similar experience of a “cold start” will be made, will feed in the preparation of the initial calibration of IRS L2 statistical retrieval.

### 3.6.2.2 Examination and correction of the surface temperature in the training set

The surface temperature ( $T_S$ ) from ERA-5 needs to be corrected to reduce some negative biases often observed in the ECMWF model data over land at daytime, especially over semi-arid regions [IASI-B L2val], before it can be used as reference data. Using the observation in a single window channel together with the PWLR<sup>3</sup> retrieved profiles and surface emissivity, it is relatively easy to detect pixels where the model  $T_S$  is too low and compute a better estimate of  $T_S$  with the help of a forward model computation. The correction  $\delta T_S$ , to be added to  $T_S$  is simply the observed minus the (clear) forward model computed radiance divided by the corresponding  $T_S$ -Jacobian as written below.

$$\delta T_S = [obs_k - F_k(T, W, O_3, T_S, \varepsilon_k)]/K_{T_S, k} \quad eq(1)$$

where

- $k$  is the channel index
- $F_k$  is the synthetic clear-sky radiance calculated with a radiative transfer model, here RTTOV
- $T, W, O_3$  are the temperature, humidity and ozone profiles, here from PWLR<sup>3</sup> retrieval
- $\varepsilon_k$  is the surface emissivity, from built-in RTTOV function over ocean and from PWLR<sup>3</sup> retrieval over land
- $K_{T_S, k} = \frac{\partial F_k}{\partial T_S}$  is the Jacobian of the forward model to the surface temperature in the window channel  $k$ .

While a positive correction  $\delta T_S$ , in the absence of inversions in the T(p) profile, can be attributed to a too low model  $T_S$ , a negative value of  $\delta T_S$  might just as well be a result of cloud contamination of the pixel as of a model  $T_S$  which is too high. Therefore the correction is only applied if it is positive and significant, i.e. above 1 K. The latter condition is in order to avoid creating a warm bias as a result of only correcting upwards (where of course random errors are expected also for the unbiased areas). Furthermore we also avoid the correction in cold areas and areas of temperature profile inversion by only adding  $\delta T_S$  to  $T_S$  if all of the four following conditions are met:

- i.  $\delta T_S > 1K$
- ii.  $IS < 1K$
- iii.  $T_S + \delta T_S > 272K$
- iv.  $T_S + \delta T_S > maxT + 1K$

where  $maxT$  is the highest air temperature over the pixel between 260 hPa and the surface and  $IS$  is the ‘inversion strength’ defined as the difference between  $maxT$  and the surface air temperature ( $T_a$ ), e.g. at 2m or at the lowermost level.

Although the surface emissivity is retrieved for the computation of the  $T_S$  correction, it is prudent to use a window channel where the variability of the emissivity is low [Capelle *et al.* 2011]. The restriction to a single channel above was done to simplify the description, in practice we compute  $\delta T_S$  for more than one channel and use the average value in order to reduce random errors. Two channels at 819.5 and 831.75 cm<sup>-1</sup> have proven to be enough in the context of IASI.

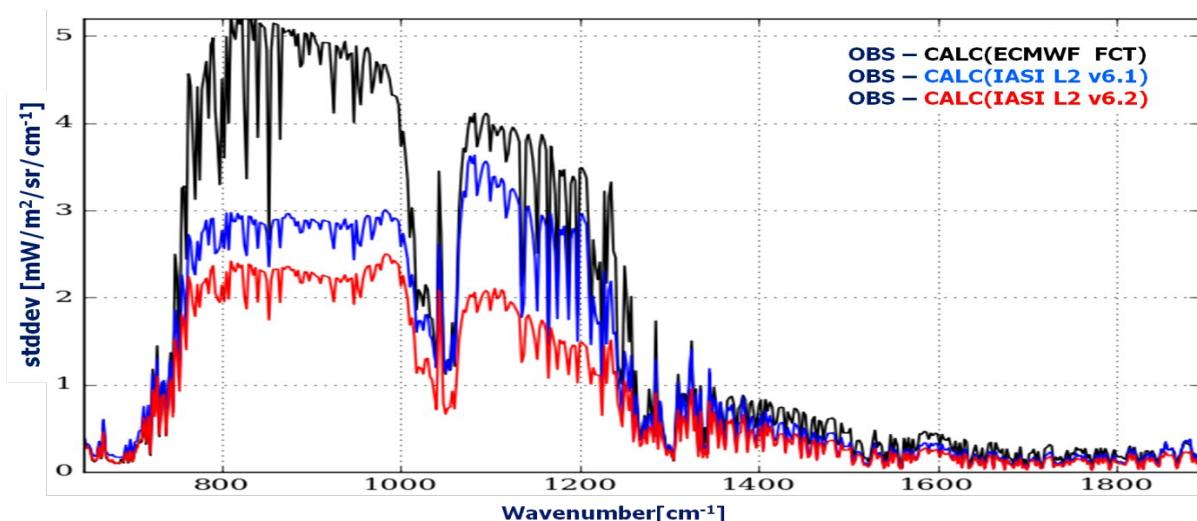
### 3.6.2.3 Definition of the cloud signal parameter

To explain the cloud signal OmC (for OBS minus CALC) to be retrieved by the PWLR<sup>3</sup>, we start by looking at a common cloud screening method in NWP and Level 2 data processing. It consists in comparing the observed window channel brightness temperature with the

corresponding brightness temperature computed by a forward model with clear-sky assumption. If the absolute value of the difference between the two is high, exceeding a configurable threshold, then the scene can be considered likely cloudy and be processed accordingly.

For the forward model computation the best available profiles shall be used and often profiles from a short range forecast are used. However, even if fine scale vertical structures of the profiles are present in the forecasts but cannot be retrieved (because they do not affect the upwelling radiance), the broader vertical structures of the profiles are usually better captured by the retrievals (with less representation error), which do therefore exhibit superior OBS minus CALC statistics in clear sky (Figure 19). Retrieved profiles are hence better suited for the detection of clouds.

In order to save online computation time during the Level 2 processing and possibly reduce random errors, the OBS minus CALC values described above are computed offline for all pairs of the training set and added to it in order to train the PWLR<sup>3</sup> for OmC along with the other geophysical parameters. As for the  $T_s$  correction (3.6.2.2) it is advisable to use an average value for two windows channels and we can use the same two centred at about 819.5 and 831.75 cm<sup>-1</sup>.



**Figure 19: Standard deviation of OBS – CALC radiances computed over continental surfaces on 16/05/2016 in the first two IASI bands. The inputs to the forward computations are collocated ECMWF forecasts + static emissivity atlas (black), PWLR (blue) and PWLR<sup>3</sup> profiles and surface parameters retrievals (red).**

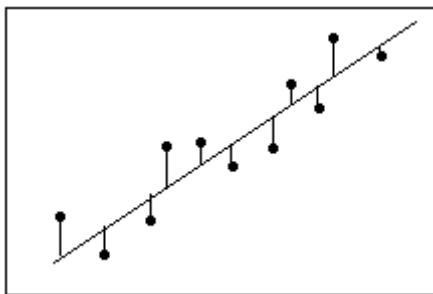
### 3.6.2.4 Computation of quality indicators

Quality indicators provide important information for the users along with the geophysical retrieved quantities supplied in the products. Depending on their applications, the users might have different selection criteria, which does not necessarily relate to the cloudiness state of the scene unlike done in previous approaches with hyperspectral missions. We propose here to complement the retrieved state parameters with individual uncertainty estimates as quality indicators.

The regression learns to estimate parameters based on a training set. A quality indicator sometimes used in regression is the static standard error of the estimates

$$\sigma_{est} = \sqrt{\frac{\sum_N (\tilde{y} - y)^2}{N}} \quad eq(2)$$

computed in a given class (3.6.1.3) with the training set.  $N$  is the training sample size,  $\tilde{y}$  and  $y$  are respectively the regressed (i.e. retrieved) and the target training state parameters in that sample.



**Figure 20: Errors between true parameters (dots) and regression estimates (line)**

The quality indicators intended for IRS are based on estimates of the errors of the retrievals. These error estimates are retrieved by the PWLR<sup>3</sup> method. To be able to do this, we need a training set of measurements and corresponding retrieval errors, for which we can use the collection of  $(\tilde{y} - y)$ , in order to apply another regression to that training set and compute the error estimates online during the retrieval stage. However, if we could estimate the errors including their sign from the PWLR<sup>3</sup> predictors, they would not be there in the first place. Instead we estimate the absolute value of the errors. The retrieval  $\tilde{y}$  of the ‘primary’ parameters  $y$  is written as

$$y \approx \tilde{y} = \bar{y} + R(x - \bar{x}) \quad eq(3)$$

with  $\bar{x}$  and  $\bar{y}$  being respectively the average predictors and dependent variables and  $R$  being the regression coefficients. From this, it follows that the retrieval of the absolute errors  $\tilde{\varepsilon}$  can be written as:

$$|\tilde{y} - y| = |\bar{y} + R(x - \bar{x}) - y| \approx \tilde{\varepsilon} = |\bar{y} + R(x - \bar{x}) - y| + R^E(x - \bar{x}) \quad eq(4)$$

where  $R^E$  are the regression coefficients for the error estimate. Hence, for each of the primary retrieved parameters we can include coefficients to retrieve the estimate of the corresponding absolute errors, which, once included in the coefficients, are retrieved in exactly the same way as the primary parameters.

An issue arises because the profiles are retrieved as PC scores (see §3.6.2.5) and it can be difficult to interpret the corresponding error estimates. Due to the fact that the error estimates apply to the absolute value, we cannot simply reconstruct the error profiles with the PCs as we do for the profiles themselves. Therefore in addition to the PC scores of the T and W profiles we also explicitly include the retrieval of the values and corresponding error estimates at selected model levels, to fulfil the need for a coarse uncertainty estimate profile expressed by Users. In this case we use the dew-point temperature for the water-vapour error estimates. . Similarly, the error estimate associated to the retrieved ozone profiles relate to the logarithm of the mixing ratio ( $\ln(\text{kg}/\text{kg})$ ) at selected levels. The table below summarises the definition of the quality indicators attached to the PWLR<sup>3</sup> retrievals.

Parameter	QI definition	Units
Temperature	Error estimates at selected levels (typically 40, configurable)	K
Humidity	Error estimates at selected levels (typically 30, configurable)	K (Dew point)
Ozone	Error estimates at selected levels (typically 10, configurable)	ln(kg/kg)
T <sub>s</sub>	Uncertainty estimate of surface skin temperature	K
Emissivity	Uncertainty estimate of the first emissivity principal component score	-

**Table 3: Definition of the PWLR<sup>3</sup> quality indicators**

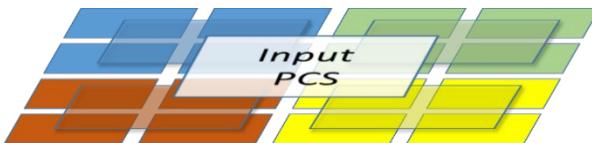
Note that while the retrieval of the primary parameters is not sensitive to random errors in the training set, provided the size of the training set is large enough, this is not the case for the retrieval of the quality indicators. Since the quality indicators are estimates of the absolute errors in the training data, the random errors in the reference data contribute to these error estimates even though they do not correspond to errors in the actual retrievals. Thus the retrieved error estimates will generally overestimate the actual errors and more so the higher the random errors in the training set are.

### 3.6.2.5 Detailed description of the retrieval algorithm

The application of the PLWR<sup>3</sup> retrieval algorithm is composed of three steps:

- the preparation of the input vector  $X$
- the computation of the output vector  $Y$  (i.e. the actual application of the piecewise linear function)
- the interpretation of the output vector.

It is a central concept of the PWLR<sup>3</sup> approach that the profiles of each individual pixel is predicted from the measurement of this pixel as well as the measurements of a group of adjacent pixels. In a conservative MTG-IRS Day-1 approach consistent with the IASI and IASI-NG schemas we choose groups of 4 by 4 pixels as the basic unit of the retrievals. This means that each input vector carries information about the IRS spectra in 16 (4 by 4) pixels and that each output vector carries information of the profiles (and other output parameters of the retrieval) for the same group of 16 pixels.



**Figure 21: PWLR<sup>3</sup> adjacent pixels grouping. Principal components of PCs in each individual IRS pixel are computed first in the 2x2-pixel quadrants (blue, green, red, yellow) and then combined again to form the inputs to the PWLR<sup>3</sup> retrieval (white box)..**

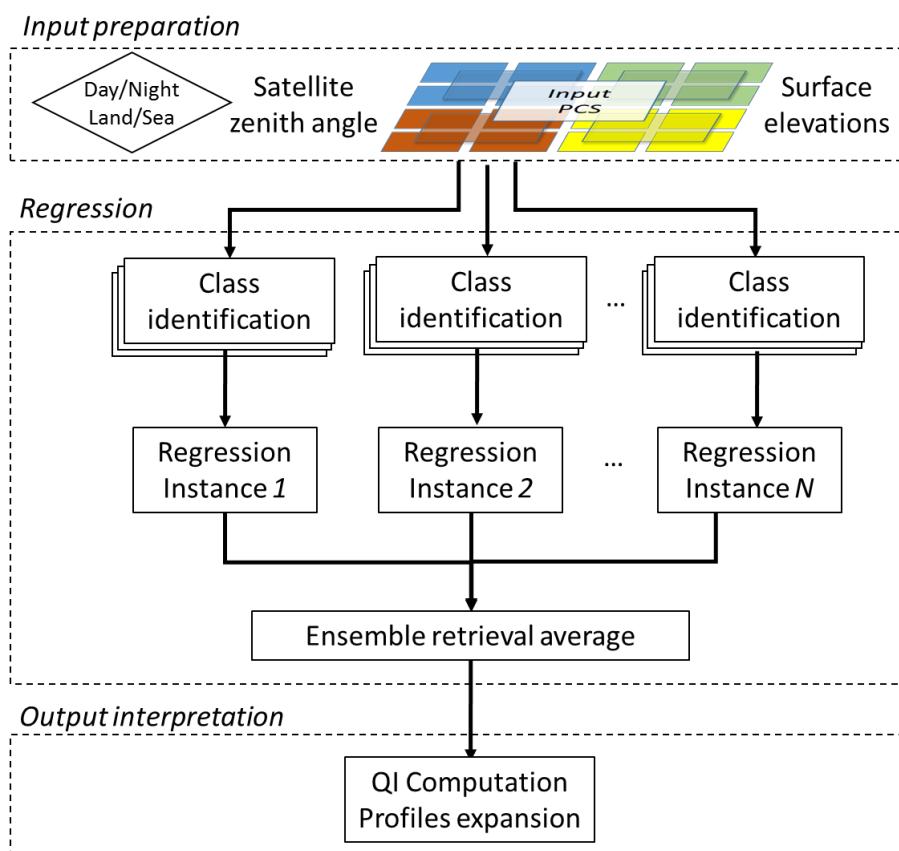
The full vector of input IRS PC scores for the 16 pixels would contain a big portion of collinearity coming both from the spectral correlation (between the two bands) and the spatial correlation (between the 16 adjacent pixels). It is therefore necessary to compute PC scores of the input PC scores. This should be done in two hierarchical steps as follows for improved computational efficiency and simpler handling of missing pixels (for example off disc pixels or pixels with bad quality spectra). First the PC scores in both band 1 and band 2 of each group of 4 adjacent (2 by 2) pixels are combined to a single set of PC scores, then the final set of PC scores for the 4 by 4 pixels is obtained by combining the 4 sets of the PC scores for 2 by 2

pixels computed in the first step. These final PC scores are supplemented in the input vector,  $X$ , by the secant of the average of the 16 satellite zenith angles and the surface elevation in meters of the 16 pixels in order. The co-linearity introduced by the inclusion of 16 adjacent surface elevations is taken care of by using ridge regression for the computation of the regression coefficients.

The input vector finally reads

$$x = (\sec \hat{z}, h_1, \dots, h_{16}, p) \quad \text{eq(5)}$$

where  $\hat{z}$  is the average satellite zenith angle,  $h_i$  is the elevation in pixel  $i$  and  $p$  are the principal component scores representing the 16 spectra in the PWLR<sup>3</sup> 4x4 group of pixels.



**Figure 22: PWLR<sup>3</sup> retrieval sequence overview**

The algorithm is the same for all four LACs, but each of the four LACs is served by a different set of PWLR<sup>3</sup> coefficients. In a given particular LAC we furthermore distinguish between four types of scenes each served by its own set of coefficients: Day-Land, Day-Sea, Night-Land and Night-Sea. If the average sun zenith angle of the 16 pixels is below 90 degrees Day coefficients are to be used (the Night coefficients are used when the average sun zenith angle is higher or equal to 90 degrees). The Sea coefficients are only to be used if all 16 pixels are over sea; if one or more of the 16 pixels are over land, the Land coefficients must be used.

Once the scene type has been determined, a number of independent retrieval instances are applied and the final retrieval vector,  $Y$ , is obtained as the average of the individual instances of retrievals. The algorithm is the same for all instances, but different sets of coefficients are used for each of them. In the following we describe the retrieval sequence, which is also illustrated in Figure 22:

- i. Regression class determination (for a single particular instance)
  - ii. Application of the regression coefficients (for a single particular instance)
  - iii. Averaging and expansion of the retrieved parameters
  - iv. Computation of the uncertainty estimates (quality indicators)
- i. **Regression class identification:** The regression class is determined by choosing the class where the centre is closest (in terms of Euclidian distance) to the scaled input vector. Only a subset of the predictors in the input vector  $X$  are used for the computation of these distances and the scale factors to be multiplied with each predictor are configuration coefficients (a scale factor equal to zero indicates that the corresponding predictor is not used for the classification and should not be used for the computation of the distances to the class centres).

$$k_x = \operatorname{argmin}_k \left( \sum_i (c_{k,i} - w_i \cdot x_i)^2 \right) \quad \text{eq(6)}$$

Where

- $k_x$  is the regression class to be used for the input vector  $x$ ,
- $k$  is the class identifier
- $i$  is the input vector index
- $c_{k,i}$  is the  $i^{\text{th}}$  element of the centre of class  $k$
- $x_i$  is the  $i^{\text{th}}$  element of the input vector
- $w_i$  is the input scaling factor of the  $i^{\text{th}}$  element of the input vector

The purpose of the input scaling factors is to avoid that the predictors with the highest variance dominate the classification, typically the standard deviation within the training set is used.

- ii. **Regression:** The class identifies which set of linear regression coefficients to be applied. It consists of mean input vector  $\bar{x}$ , mean output vector  $\bar{y}$  and regression operator  $R$ . A separate set of linear regression coefficients is available for each of the classes. The application of these coefficients produces the instance output vector  $y$  as

$$y = \bar{y} + R(X - \bar{x}) \quad \text{eq(7)}$$

The output vector computed as described above represents the retrieved parameters of the 16 individual pixels, where T, W, O and E are represented as PC scores. It is possible to have individual PCs for each of the 16 pixels or to have common PCs which cover the profiles (or emissivity spectra) of the 16 pixels. Both strategies work and have been demonstrated with IASI. The former approach is retained here, so the retrieved vector writes.

$$y = (s_1^T, \dots, s_{16}^T, s_1^W, \dots, s_{16}^W, s_1^O, \dots, s_{16}^O, s_1^{CO2}, \dots, s_{16}^{CO2}, Ts_1, \dots, Ts_{16}, \\ s_1^{\text{emiss}}, \dots, s_{16}^{\text{emiss}}, OmC_1, \dots, OmC_{16}, \dots, \\ QI_1^T, \dots, QI_{16}^T, QI_1^W, \dots, QI_{16}^W, QI_1^O, \dots, QI_{16}^O, QI_1^{Ts}, \dots, QI_{16}^{Ts}, QI_1^{Ps}, \dots, QI_{16}^{Ps}, QI_1^{\text{emiss}}, \dots, QI_{16}^{\text{emiss}}) \quad \text{eq(8)}$$

Where  $s_i^x$  are the principal component scores (vectors) and  $QI_i^x$  are the quality indicators of parameter  $x$  (profiles/vectors in the case of T, W, O<sub>3</sub>, CO<sub>2</sub> and emissivity) in pixel  $i$ .

- iii. **Output interpretation:** The geophysical parameters retrieved with the different instances of the PWLR<sup>3</sup> are averaged at this stage. For the parameters represented in principal components, they are averaged before being expanded, e.g. to vertical profiles or surface emissivity spectra. The expansion writes as follows:

$$P_k = [\bar{P}_k + \sum_{i=1}^n E_{k,i} s_i] \quad \text{eq(9)}$$

where  $P_k$  is the retrieved quantity (i.e. temperature, humidity, ozone or emissivity) at index (i.e. level or channel)  $k$ ;  $\bar{P}_k$  is its mean,  $E$  is the eigenvectors for this parameter and  $s$  the corresponding principal components scores  $i = 1..n$ . Indicatively 40, 30, 20 and 10 principal components are intended for Temperature, Humidity, Ozone and Emissivity. In the implementation of the operational code, these dimensions shall be configurable.

The profiles retrieved by PWLR<sup>3</sup>, after expansion of the PC scores, are sampled at hybrid model levels known from ECMWF. This means that the surface pressure, which is also retrieved by PWLR<sup>3</sup>, is needed to associate pressure levels for each profile as described in [ECMWFgrid137]. When the pressure levels are known the profiles can be interpolated as described in 4.1 to the standard 101 level pressure grid used for the output product (and for the RTTOV coefficients). It is planned to represent water-vapour and ozone profiles in the logarithm of the mixing ratio within the PWLR<sup>3</sup> regression. Then the profiles are converted into SI concentration units in the final products.

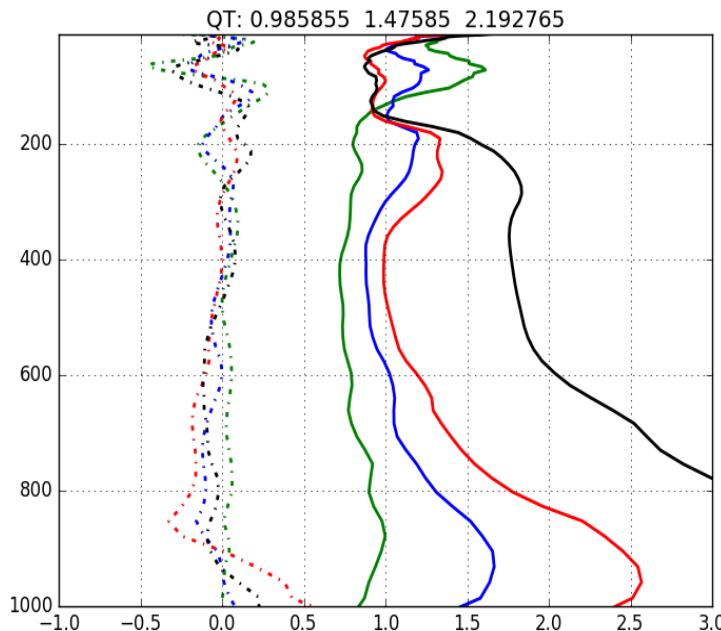
- iv. **Quality indicators:** We recall that the quality indicators are obtained by retrieval of the absolute error of the parameter in question. Like the other retrieved parameters, the quality indicators are retrieved by multiple retrieval instances, but unlike the other parameters we do not always use the average of the retrieved instances as the final retrieval of the quality indicators. Instead we use the standard deviation (over the instance retrievals) of the underlying parameter to which the quality indicator applies, whenever this standard deviation is greater than the usual average of the retrieved quality indicators. The standard deviation of the underlying parameter can be thought of as an estimate of the retrieval noise which is just a part of the total retrieval error estimated by the quality indicator and is therefore expected to be lower. However since there is uncertainty in the estimation of the absolute error, it can (and does) happen that the error estimate (of the total error) is lower than the estimate of the retrieval noise alone, which justifies the choice to use the maximum of the two as the final quality indicator.

$$QI_j = \max \left( \text{stddev}_i(\tilde{y}_{i,j}), \text{mean}_i(\tilde{\varepsilon}_{i,j}) \right) \quad \text{eq(10)}$$

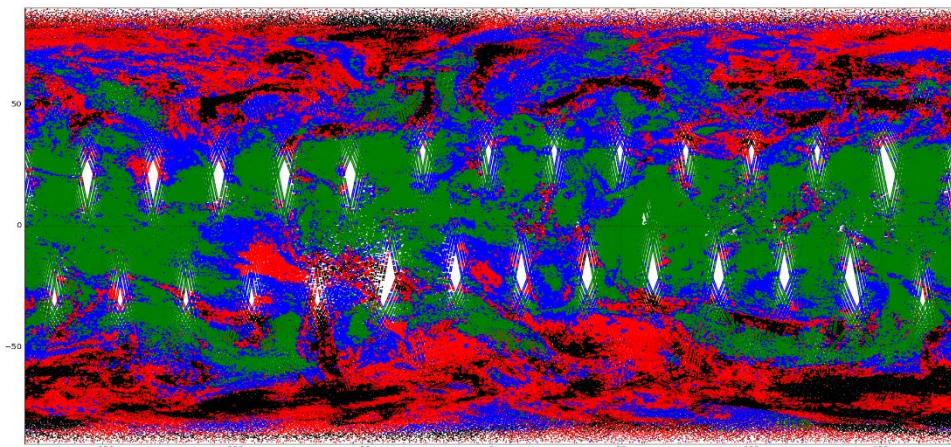
Where  $\tilde{y}$  is the retrieved parameter and  $\tilde{\varepsilon}$  is the predicted uncertainty estimate computed after eq(4).  $i$  is the retrieval instance and  $j$  is the geophysical parameter identifier (e.g. Temperature at a given pressure level, or skin surface temperature...).

Figure 23 shows statistics of IASI PWLR<sup>3</sup> temperature retrievals vs ERA-5 reanalysis on 04/10/2016. The global dataset of all sky retrievals covering all pixels for the full day were partitioned in 4 quality classes of equal size (i.e. 25% of all pixels each) based on a unique temperature quality indicators corresponding to low tropospheric levels. The thresholds on the temperature quality indicators (error estimates) used for the partitioning were approximately 1 K, 1.5 K and 2.2 K. We observe that the retrievals in the best quality class, depicted in green, are as expected the most precise (below 200 hPa) (as evaluated against ERA-5). The statistics of the second best quality class is depicted in blue, the third best in red and the worst in black – the reliability of the temperature quality indicators is clearly demonstrated. Figure 24 shows the geographical distribution of the quality classes on the same day. The colours used are the same as in the previous figure – green, blue, red, black in order of decreasing quality. As infrared radiances are not sensitive to the atmosphere below clouds there is a high degree of correlation between the cloudiness and the quality indicators. These results are provided to

illustrate the significance of such retrieval estimates (here a single scalar quality indicator). In the operational implementation, the algorithm will generate uncertainty profiles for the atmospheric parameters, and an uncertainty vector for the emissivity. The units should be configurable. Good experience has been made so far of expressing uncertainties in K for temperature, dew point K for humidity and logarithmic mixing ratio for ozone.



*Figure 23: Stratification of Temperature statistics (dash: bias, solid: standard deviation) IASI-PWLR<sup>3</sup> vs ECMWF analysis on 04/10/2016, partitioning in 4 quality classes of equal size based on quality indicator.*

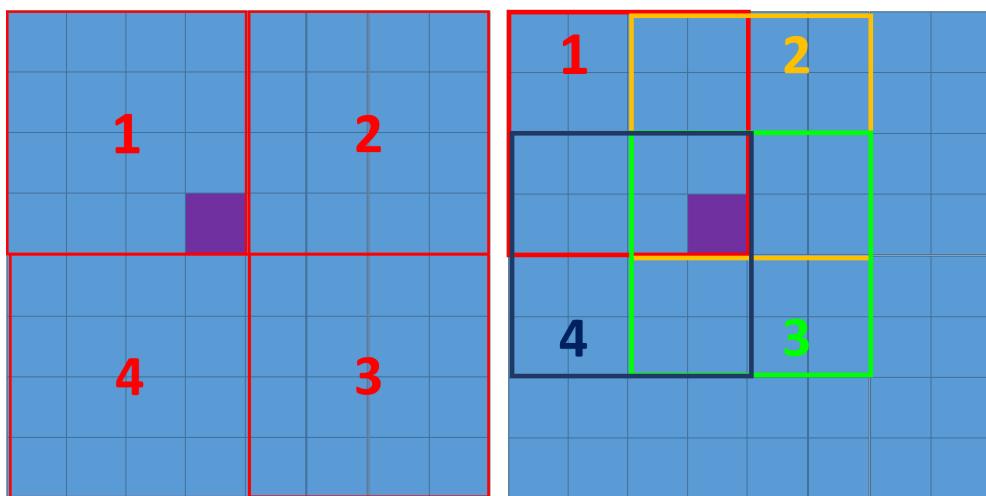


*Figure 24: Map of the quality classes defined above.*

### 3.6.2.6 Moving window strategy

If the PWLR<sup>3</sup> 4x4 window was moved in a contiguous manner, as illustrated in Figure 25 left-hand side, each dwell would be covered by 40 times 40 PWLR<sup>3</sup> retrievals of 4 by 4 groups of pixels. With this configuration, pixels like the lower-right corner highlighted in purple in Figure 25 would only benefit from neighbouring pixels from one side. To avoid this, the PWLR<sup>3</sup> window could be translated by one pixel at a time and retrievals be retained only for the centre

pixel, which would effectively result in 25600 retrievals per dwell. To mitigate the computation overhead and still ensure that retrievals in individual pixels will benefit from adjacent information in all directions, overlapping groups of 4 by 4 retrievals will be organised, yielding more than one PWLR<sup>3</sup> retrieval for each pixel. For Day-1, we suggest to use overlapping 4 by 4 groups always shifted by two pixels in each direction, as illustrated in Figure 25 right-hand side. In this way we end up with 79 times 79 basic PWLR<sup>3</sup> retrievals per dwell and all pixels, except at the edges of the dwell, will be a combination of 4 different retrievals. Of these four retrievals there will always be one in which the current pixel is one of the four centre pixels in the 4 by 4 group (green-window 3 in Figure 25 right-hand side), there will be two where it is on the lateral edge but not corner (yellow- and blue-windows 2 and 4) of the 4 by 4 group and one where it is in the corner (red-window 1). The four individual retrievals for the current pixel as detailed in 3.6.2.5 will be averaged with weights 1/3, 1/4, 1/4 and 1/6 respectively to form the final retrieval for this pixel, to give more weight to the configuration where it sits in the middle of the window.



*Figure 25: PWLR<sup>3</sup> moving window strategy illustrated on an 8 by 8 IRS-pixels subgrid of a dwell. Left: non-overlapping groups. Right: overlapping groups*

### 3.6.3 Handling non-nominal situations

The regression assumes that all predictors, with which it was trained, are available and of good quality. This can cause problems, when one or more of the 16 adjacent pixels used in a PWLR<sup>3</sup> retrieval is missing or of bad quality. To overcome this problem we can use the remaining good pixels to predict the failed or missing pixels with linear regression and compute the PC scores of the predictors with predicted values of the missing predictors, which can be achieved with a simple update of the eigenvectors.

Let  $E$  be the matrix of eigenvectors, partitioned into rows corresponding to missing predictors  $E_1$  and rows corresponding to the remaining good predictors  $E_0$

$$E = \begin{bmatrix} E_0 \\ E_1 \end{bmatrix} \quad \text{eq(11)}$$

and let  $C$

$$C = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix} \quad \text{eq(12)}$$

be the covariance matrix of all predictors compatibly partitioned.

To eliminate the use of the missing predictors,  $E_1$  must be set to zero in the updated eigenvectors and  $E_0$  must be replaced by

$$E_0 + C_{00}^{-1}C_{01}E_1 \quad \text{eq(13)}$$

Using this principle, a total of 15 different sets of eigenvectors are needed for the computation of the PC score for groups of 2 by 2 pixels to cover all possible combinations of zero to three bad or missing pixels out of the four. Likewise, for the second step combining four groups of 2 by 2 into PC scores for a group of 4 by 4, we need 15 different set of eigenvectors to be able to handle all possible combinations of missing 2 by 2 groups (except all four subgroups missing of course).

### 3.7 Cloudy scene characterisation

#### 3.7.1 Cloud retrieval

##### 3.7.1.1 Retrieval principle

The purpose of this algorithm is to identify and characterise the presence of clouds in the field of view. A simple but useful model in satellite products represents the clouds as black bodies, characterised with an emitting top located at a certain altitude (pressure level) and an effective fractional coverage [Eyre and Menzel 1989]. The effective fractional coverage accounts for the geometrical extent as well as for the transparency of the cloud (e.g. a semi-transparent cloud with full extent over the pixel will have an effective coverage of 50%). Describing the radiative transfer within a cloud and aiming at more exhaustive description of it, e.g. including vertical profiles and microphysics parameters, is still a subject of active research in the thermal infrared. The computation-time of such models is also not compatible with near-real time operational applications yet.

With this simple cloud modelling, the observed radiances at the top of the atmosphere  $R_\nu$  at wavenumber  $\nu$  writes as a linear combination of the clear-sky and cloudy radiance terms:

$$R_\nu = (1 - \alpha_\nu) * R_\nu^{clear} + \alpha_\nu * R_\nu^{cloudy}(p) \quad \text{eq(14)}$$

where  $R_\nu^{clear}$  is the cloud-free radiance,  $R_\nu^{cloudy}(p)$  the overcast radiance with cloud-top at pressure  $p$  and  $\alpha_\nu$  is the effective cloud coverage at wavenumber  $\nu$ .

When using forward models like RTTOV [RTTOV 12] to simulate clear-sky top of the atmosphere radiances, the simulated overcast radiances with a hypothetical cloud top height at several different pressure levels are generally obtained as a by-product with no additional computational cost. This has been used to retrieve the cloud top height by fitting the measured radiances in selected channels by varying the cloud top height and the effective cloud fraction in the resulting simple cloudy forward model [Stubenrauch *et al.* 1999].

Other cloud top height retrieval schemes based on this simple cloud model, which have been proposed for hyperspectral infrared data, include CO<sub>2</sub> slicing [Smith and Frey 1990] and MLEV [Huang *et al.* 2004]. Essentially these methods are similar to the radiance fitting method, but introduce a dependency of the channel weights on the cloud top pressure. As scattering is not taken into account in the simple (and fast) cloud model we only use channels between 700 and

950 cm<sup>-1</sup> where the effect of scattering is small [Huang et al. 2004]. The latter two methods get very sensitive to input uncertainties when the cloud signal (difference between clear-sky and overcast radiances components) is small as it is used as scaling factor in the denominator of the cost function to be minimised.

The algorithm described hereafter builds on the method in [Stubenrauch et al. 1999] and extends it to investigate potential 2-layer cloud configurations. Likewise, it assumes that the effective coverage is independent of the wavenumber.

### 3.7.1.2 Computing the costs with clear and cloudy assumptions

Let

- $\tilde{r}_k$  be the reconstructed radiances measured by MTG-IRS
- $r_k$  the bias corrected radiance  $r_k = \tilde{r}_k - b_k$ , with  $b_k$  the bias applied in channel  $k$
- $y_k$  be the simulated clear radiances (see section 3.7.1.4 for RTM and input vectors),
- $c_{jk}$  be the simulated overcast radiances for a hypothetical cloud at pressure level  $j$  (see section 3.7.1.4 for RTM and input vectors)
- $S_y$  the observation error covariance ( $k \times k$ ) matrix for the channels involved in the cloud retrieval.

where  $k$  is the channel index running over the channels used in the simple cloud parameter retrieval.

The cost to be minimized generically writes:

$$C = (r - F)^T S_y^{-1} (r - F) \quad \text{eq(15)}$$

Where  $F_k$  is the forward-modelled radiance (with RTTOV) computed as follows under clear and cloudy assumptions.

#### Clear-sky assumption:

The cost  $C_0$ , or weighted radiance misfit, with the clear-sky assumption writes from  $C$  with:

$$F_k = y_k + d_{T_S} K_{T_S, k} \quad \text{eq(16)}$$

To allow for some uncertainty in the surface skin temperature, the above simple clear sky cost is evaluated having fitted the surface skin temperature ( $T_S$ ), subject to a maximum absolute difference from the initial  $T_S$  of 1 K.

The surface temperature increment,  $dT_S$ , can be computed as follows:

$$dT_S = \min \left\{ 1, \max \left\{ -1, \frac{K_{T_S}^T S_y^{-1} (r - y)}{K_{T_S}^T S_y^{-1} K_{T_S}} \right\} \right\} \quad \text{eq(17)}$$

Where  $K_{T_S, k}$  is the derivative of the forward model with respect to surface temperature at channel  $k$ .

#### One-cloud assumption:

For a one-layer cloud at pressure level  $j$  with effective cloud fraction  $\alpha$ , the cloudy cost  $C_1$  derives from  $C$  with

$$( )$$

$$F_k = (1 - \alpha)y_k + \alpha c_{jk} \quad \text{eq( 18)}$$

which must be minimized by varying  $j$  and  $\alpha$ . The weights and correlations in the  $S_y$  matrix are configurable and account for the overall uncertainties in fitting the observations with synthetic calculations at the different channels. Like in OEM, they include instrument noise and errors in the forward modelling and can be evaluated empirically. For any fixed cloud top pressure at level  $j$ , the minimization problem has the analytical solution

$$\alpha_j = \min \left\{ 1, \frac{(c_j - y )^T S_y^{-1} (r - y )}{(c_j - y )^T S_y^{-1} (c_j - y )} \right\} \quad \text{eq( 19)}$$

Levels where  $\alpha_j$  is zero or negative are not considered further, since this does not correspond to possible cloud. A cloud fraction  $\alpha_j$  greater than 1 does not correspond to any possible cloud either, but for such levels we can consider a cloud fraction equal to 1 and compute the corresponding value of the cost function. The retrieved (one layer) cloud top pressure is the level for which the cost function yields the lowest value; only considering pressure levels below the tropopause, or below 100 hPa if it cannot be determined [Reichler et al. 2003]. Reichler's approach follows the [WMO 1957] definition, where the tropopause is defined as ,“the lowest level at which the lapse-rate decreases to 2°C/km, or less, provided that the average lapse-rate between this level, and all higher levels within 2 km does not exceed 2°C/km”.

### **Two-cloud assumption:**

Retrieving two-layer clouds is envisaged in future, with further studies to evaluate the potential. We provide below the theoretical mathematical formulation. This will not be implemented for Day-1.

In situations where two different cloud top levels exist within the field of view, a better radiance fit might be obtained when using a two-layer cloud model:

$$C_2(i, j, \alpha^1, \alpha^2) = (r - (1 - \alpha^1 - \alpha^2)y - \alpha^1 c_i - \alpha^2 c_j)^T S_y^{-1} (r - (1 - \alpha^1 - \alpha^2)y - \alpha^1 c_i - \alpha^2 c_j) \quad \text{eq( 20)}$$

Similarly to the one-layer cloud model, the optimal cloud fractions can be determined analytically for any fixed pair of cloud levels  $i < j$

$$\alpha_{ij}^1 = \frac{(r - y - A(c_i - y ))^T S_y^{-1} (c_j - y + B(c_j - y ))}{(c_i - y + B(c_j - y ))^T S_y^{-1} (c_i - y + B(c_j - y ))} \quad \text{eq( 21)}$$

$$\alpha_{ij}^2 = A - B\alpha_{ij}^1 \quad \text{eq( 22)}$$

where

$$A = \frac{(c_i - y )^T S_y^{-1} (r - y )}{(c_j - y )^T S_y^{-1} (c_j - y )} \quad \text{eq( 23)}$$

and

$$B = \frac{(c_j - y )^T S_y^{-1} (c_i - y )}{(c_j - y )^T S_y^{-1} (c_j - y )} \quad \text{eq( 24)}$$

If a pair of cloud levels  $i < j$  exists such that the corresponding two-layer cloud model cost function value is lower than the best one-layer cost, then the two-cloud layers representation is retained, with cloud top pressures found by choosing the pair  $i < j$  which has the lowest cost. In this process, similarly to the one-layer clouds, if any of the individual cloud fractions  $\alpha_{ij}^1$  and  $\alpha_{ij}^2$  or the sum of the two is greater than 1 we must of course evaluate the cost after normalising the cloud fractions so that their sum is 100%, the natural limit. Like in the 1-layer cloud formation, only pressure levels below the tropopause are considered, or below 100 hPa if it cannot be determined [Reichler et al. 2003].

### 3.7.1.3 Clear/Cloudy costs assessment and scene classification

The best clear and cloudy costs as described in §3.7.1.2 are compared, using their ratio

$$q = \frac{\min(C_1, C_2)}{C_0} \quad \text{eq( 25)}$$

If the cloud-free cost is smaller than the best cloudy cost, then the scene is declared ‘cloud-free’. If the best cloudy cost (1- or 2-layer) is lower than the clear-sky but the relative improvement is too small as assessed with a configurable threshold  $\theta_0^{cld}$

$$1 \geq q \geq \theta_0^{cld} \quad \text{eq( 26)}$$

then the detection of a cloud is considered ambiguous. Practically, the scene is classified as ‘clear enough’ and the cloud parameters (fraction and top height) are unset. Physically, this would correspond to situations where the clear-sky assumption can be considered valid for simulating IRS observations, within the uncertainties of the forward modelling due to the RTM itself or to the input parameters.

The resulting scene classification is recorded in a processing flag, the cloudiness flag, for information to the users and as input to subsequent processing functions. Such pixels classified as ‘cloud-free’ or ‘clear enough’ can be further processed with the optimal estimation described in the section 3.8, which assumes clear-sky. A cloud is flagged when the cloudy cost is significantly smaller in relative terms than the clear cost as assessed with  $\theta_0^{cld}$ . The associated cloud fraction(s) and cloud top height(s) are then reported in the final product. The value of  $\theta_0^{cld}$  for IRS is expected to be similar to that used for IASI: which is 0.5. We distinguish the intensity of the cloud contamination with a second threshold  $\theta_1^{cld}$ , set to 0.05 for IASI.

Cloudiness summary		
Value	Meaning	Reason for flagging
1	Clear-sky	Cloud-free cost $C_0$ is the lowest
2	Clear enough, some small cloud contamination possible	The relative improvement of the cost with clouds is small $1 > q \geq \theta_0^{cld}$
3	Cloudy – moderate cloud signal	Cloudy cost $C_1$ or $C_2$ is the lowest and $\theta_0^{cld} > q \geq \theta_1^{cld}$
4	Cloudy – strong cloud signal	$q < \theta_1^{cld}$

*Table 4: Definition of the cloudiness summary*

### 3.7.1.4 Input profiles to RTTOV simulations for cloud retrievals

The profiles used in the invocation of RTTOV could be taken from either the PWLR<sup>3</sup> retrieval or from the numerical forecasts; the choice should be configurable. In the context of IASI, the use of PWLR<sup>3</sup> profiles has been successfully tested, but for IASI the PWLR<sup>3</sup> nominally does also make use of collocated microwave (AMSU/MHS) observations, which improves the retrievals below clouds. Companion microwave measurements will not available in the context of IRS. Detecting clouds in the lower troposphere with forecast fields is particularly challenging over land due to uncertainties in the forecasts, assumptions made on the static state vector (e.g. emissivity) and uncertainties in the radiative transfer modelling [McNally and Watts 2003]. OBS minus CALC statistics computed with PWLR<sup>3</sup> in clear sky exhibits smaller standard deviations than when using forecast profiles in general, and particularly over continental surfaces [IASI L2 v6.4 val]. The PWLR<sup>3</sup> profiles hence represent an interesting alternative input to distinguish small cloud contaminations from clear sky situations. With higher cloud amount, the errors of the infrared only PWLR<sup>3</sup> retrievals gets higher, which is balanced by the fact that the contribution of the part of the profiles below the cloud top gets lower. The actual choice of the input profiles will be reflected in a processing flag in the final product.

### 3.7.1.5 Channel selection and weighting for cloud retrievals

The channel selection should be done such that the total information content in the spectral region we consider is preserved, following a similar method as presented for the channel selection of the optimal estimation (see §3.8.3.2). In theory, bias correction should be applied to the observations to take any systematic differences with respect to the model into account, but since this is hard to obtain for cloudy radiances it is not foreseen at this stage. In the current formulation, we have not included inter channel correlations in the cost functions, which are parameterized just by the channel weights,  $w_k$ . In most scenes, the cloud signal is sufficiently unambiguous, such that the retrieved cloud top heights are relatively insensitive to the weights. For the remaining scenes it might be beneficial to change to a full matrix formulation of the cloud cost functions and determine a weight matrix  $W$  based on a similar approach to how the observation error covariance matrix  $S_y$  is determined for the optimal estimation (see §3.8).

## 3.7.2 Cloud phase

The distinction between ice and liquid clouds can be done from their emission in the thermal infrared with simple brightness temperature difference tests [Strabala and Menzel 1994]. The test described below exploits a spectral region where the refractive indices of water and ice are distinct (between 11 and 12  $\mu\text{m}$ ) and another region where they are comparable (between 8 and 11  $\mu\text{m}$ ). Corresponding channels can be selected in the infrared atmospheric windows where relatively less absorption is occurring (especially of water-vapour), so that brightness temperature differences essentially come from the different cloud characteristics [Chylek *et al.* 2006].

The discrimination of ice from water clouds makes use of the brightness temperatures  $T_b$  at 8, 11 and 12  $\mu\text{m}$ .

If

$$\{T_b(8\mu\text{m}) - T_b(11\mu\text{m})\} - \{T_b(11\mu\text{m}) - T_b(12\mu\text{m})\} > \theta_1 \quad \text{eq(27)}$$

then the cloud phase is ice, else if

$$\theta_1 > \{T_b(8\mu\text{m}) - T_b(11\mu\text{m})\} - \{T_b(11\mu\text{m}) - T_b(12\mu\text{m})\} > \theta_2 \quad \text{eq(28)}$$

then the cloud is of mixed phase. The cloud is of water phase otherwise.

$\theta_1$  and  $\theta_2$  are configurable thresholds. As an indication, they are set to -0.9 and -1.15 in the IASI operational processor.

A possible evolution would involve a statistical cloud classification/recognition algorithm, with supervised learning including for instance with collocated infrared measurements and cloud masks from high spatial resolution imagers.

### 3.7.3 Dust detection

Windblown dust and aerosols have low frequency spectral signature in the thermal infrared, with dust yielding a W-shape depression at first order in the 800-1200 cm<sup>-1</sup> region [Sokolik 2002] [De Souza *et al.* 2006]. The intensity of the aerosol signal is related to the load but also to the type and size of particles. Other particles, like ice crystals, sulfate droplets or biomass burning aerosols yield distinct low frequency spectral signal, of different slope signs and strengths [Clarisso *et al.* 2013]. These signatures have been long exploited for the detection and characterisation of aerosols, including a number of brightness temperature difference tests from the legacy broadband imager and infrared sounding missions.

The methodology described hereafter is a simple implementation of [Clarisso *et al.* 2013], taking advantage of the high spectral resolution of hyperspectral sounder. It aims at retrieving a pseudo-quantitative unitless indicator of the dust load in a given field of view. This dust indicator has been demonstrated with IASI and significant correlations with AOD retrieved from other sources were observed [Clarisso *et al.* 2013]. This indicator also correlates well with the loss of accuracy in SST retrievals when aerosol are not taken into account [Trent *et al.* 2016]. Preliminary studies have shown that the methodology should be successfully applicable to IRS resolution as well, because of the slow-varying spectral signature of dust and given that the 800-1200 cm<sup>-1</sup> region is well covered by MTG-IRS.

The retrieved dust indicator with this approach is a weighted projection of the observed signal onto typical aerosol signature:

$$R = k^T S^{-1} (y - \mu_c) \quad \text{eq( 29)}$$

where:

- $k$  is the dust Jacobians
- $S$  is the covariance matrix of dust-free spectra
- $y$  is the observed spectrum
- $\mu_c$  is the mean dust-free spectrum.

$k$ ,  $S$  and  $\mu_c$  can be defined empirically from a static training base of real observations, including dust-free and dust-contaminated pixels. The training base can be stratified with classification techniques such as k-mean.  $\mu_c$  and  $S$  are then the average spectrum and covariance within the dust-free class. As an alternative to explicit forward modelling,  $k$  is approximated as the difference  $\mu_p - \mu_c$  where  $\mu_p$  is the mean spectrum in the class of polluted/dust-loaded spectra. Different static configuration of the gain matrix  $G = k^T S^{-1}$  are pre-computed off-line for the different land/sea and day/night combination separately.

For clear spectra, this quantity has a mean of zero and a standard deviation of one. The presence of dust can be suspected when the indicator  $R$  exceeds a configurable threshold, typically of 2 to 3.

Further evolution could include the determination of different aerosol types and ice particles with classification algorithms.

### 3.8 Second retrieval: Optimal Estimation Method

#### 3.8.1 Overview

While the first retrieval is performed in both clear and cloudy situations, the second retrieval – the optimal estimation method – is only performed in pixels which are clear, or considered clear enough to apply cloud-free radiative transfer modelling. Optimal estimation is a variational method formulated as a minimization problem, seeking the state vector for which the application of a forward model simulation yields radiances which best match the observations and a regularization term which penalizes deviations of the state vector from the so called *a priori* state, as described in great detail in [Rodgers 2000]. As a baseline, RTTOV 12.1 will be used as the forward model, but it is recommended to make the implementation in such a way that the forward model can be exchanged without impacts on the other parts of the retrieval code.

As mentioned above, the cost function consists of two terms, which we will refer to as the background (synonym for *a priori*) and the observation term. Besides the choice of *a priori* and forward model, the optimal estimation method (OEM) is determined by the representation of the state-vector and the observations as well as the background error covariance matrix  $S_x$  and the observation error covariance matrix  $S_y$ .

$$J = J_x + J_y = (x - x_a)^T S_x^{-1} (x - x_a) + (F(x) - y)^T S_y^{-1} (F(x) - y) \quad \text{eq(30)}$$

#### 3.8.2 Background term of the cost function, $J_x$

$$J_x(x) = (x - x_a)^T S_x^{-1} (x - x_a) \quad \text{eq(31)}$$

The state-vector  $x \in \mathbb{R}^n$  represents the quantities to be retrieved by the optimal estimation method:

- temperature profile (T)
- water vapour profile, specific humidity (W)
- ozone profile (O)
- CO<sub>2</sub> profile CO<sub>2</sub> (optional)
- surface skin temperature ( $T_s$ )

The profiles are represented as principal component (PC) scores (as further explained in the forward model section) of the deviation with respect to the *a priori*. The number of principal component scores used for each of the three types of profiles must be configurable – typical values might be  $n_T = 28, n_W = 18, n_O = 10, n_{CO_2} = 3$ . The total number of state-vector elements is  $n = n_T + n_W + n_O + n_{CO_2} + 1$ .

The PC scores used in the state-vector representation of the T, W, O and CO<sub>2</sub> profiles are based on deviations from the *a priori* profile given at the 101 fixed pressure levels RTTOV grid. The units of the profiles on which the PC's are based are T: K, W: ppmv, O: ppmv and C: ppmv using the logarithmic values in the case of W and O. It is possible to exclude one (or more) of

the four profiles from the active part of the state-vector by configuring the corresponding number of PC scores to be zero. Note: at the time of writing this version of the document, CO<sub>2</sub> is not yet an active (i.e. being retrieved) state vector parameter in the IASI operational processor. This corresponds to a configuration with  $n_{CO_2} = 0$ , in which case the CO<sub>2</sub> profile obtained in the first retrieval is carried over as input to the forward RTM of the OEM and is not allowed to change.

### 3.8.2.1 Choice of *a priori*, $x_a$

The *a priori*  $x_a$  could either be taken from the first (PWLR<sup>3</sup>, section 3.6) retrieval or from collocated forecast fields (section 3.5.4). In terms of implementation, this choice is irrelevant and should be configurable for real-time operations.

We have shown that the measurements can be better fitted when spectra are simulated with the PWLR<sup>3</sup> profiles than with collocated forecast profiles [IASI L2 v6 val][IASI L2 v6.2 val][IASI L2 v6.4 val]. Using the PWLR<sup>3</sup> profiles as first guess too is therefore expected to minimise the number of steps required to reach convergence in the OEM. It is understood that using the first retrieval as *prior* deviates from the theoretical definition of an *a priori* in the strict sense of Rodgers, i.e. being an independent information before the observation is made. However, being the most likely atmospheric state inferred from the observations based on an exhaustive climatology (that of ECMWF analyses), the first retrieval has proven to be a very valuable regularisation point in the minimisation process. Furthermore, this configuration would have the potential advantage that the retrievals will be independent of the corresponding NWP forecast data, which are already available to the forecasters.

On the other hand, high frequency structures may be contained in the forecast profiles which cannot be observed by the MTG-IRS instrument because they lie in the so-called null space, i.e. they do not have any impact on the measured (nor simulated) radiances. These high frequency perturbations could have an interest for regional applications, e.g. to determine the boundary layer for instance or low level smaller inversions [Eyre et al. 2011]. They would be included in the profiles resulting from the optimal estimation if numerical predictions containing these signals were used as *a priori*.

Diverging requirements have been expressed by Users with respect to the dependency to forecasts [IRS MAG 5], coming from different application –e.g. AMVs need be derived without forecasts prior if to be assimilated in models afterwards- but also within the Nowcasting community. There, it was expressed that independent observations -possibly at coarser vertical resolution- could suit the needs of forecasters. The release of the regional service EARS IASI L2 in a pilot phase in November 2017 [EARS IASI L2] allowed close interactions with end Users to study and discuss these matters. The potential of hyperspectral sounding products for regional applications is now being investigated in cooperation with NMHSs, the Convection Working Group and the European Severe Storm Lab, and through dedicated studies e.g. [Kocsis et al . 2017], bearing MTG-IRS objectives and specificities in mind.

The choice of the *a priori* is configurable and will be reflected in a processing flag of the final products. Possible fine adjustments to the algorithms can be made incrementally from this basis, including lessons from testing IASI and proxy-IRS L2 products with the Users and iterating on their requirements. The operational retrieval operator for IASI, which this IRS

ATBD inherits from, has been successfully tested with one or the other option [Crapeau *et al.* 2017].

We note that, no matter what we choose as *a priori*, the actual vector,  $x_a$ , is zero because we represent the state vector as PC scores of the deviations with respect to the chosen *a priori* profiles.

### **3.8.2.2 *A priori* error covariance matrix, $S_x$**

The eigenvectors used for the state-vector representation are computed offline along with the corresponding *a priori* error covariance matrix,  $S_x$ . This computation is based on the covariance matrix of the differences between the *a priori* profiles and the collocated NWP analysis profiles. In the option using numerical forecasts as *a priori*, the background covariance matrix defined for Day-1 will be computed on a static climatology of forecasts vs analyses. A dynamic background matrix reflecting more accurately the statistics of the forecast error, for instance using the ensemble-based covariance at ECMWF, will be investigated as part of Day-2 evolutions (OI-8).

A separate covariance matrix is computed for each of the geophysical parameter to be retrieved. In the PC space in which the profiles are retrieved, the background error covariance matrix is diagonal and composed of four submatrices – one for each type of profile – on the diagonal and a last diagonal element corresponding to the first retrieval error variance of the surface skin temperature.

As the PC scores used in the state-vector representation of the profiles are based on deviations from an *a priori* profile given at the 101 fixed pressure levels RTTOV grid, the *a-priori*  $x_a$  in this representation is zero (except of course for the last element corresponding to  $T_S$ ).

### **3.8.3 Observation term of the cost function, $J_y$**

$$J_y(x) = (F(x) - y)^T S_y^{-1} (F(x) - y) \quad \text{eq(32)}$$

where

- $x$  is the state vector to be retrieved
- $y$  is the observation vector, i.e. reconstructed IRS radiances in selected channels
- $F(x)$  is the forward model function
- $S_y$  is the observation error covariance matrix, including instrument and forward modelling errors

#### **3.8.3.1 Forward model, $F$**

The underlying forward model is RTTOV 12.1 using the rttov8pred101L coefficients (i.e. the coefficients for 101 levels pressure grid, with variable O<sub>3</sub> and CO<sub>2</sub>, but fixed CH<sub>4</sub>, CO or N<sub>2</sub>O) [RTTOV 12]. The state-vector in the optimal estimation is represented as principal component scores of the deviation with respect to the *a priori* state vector. This has the advantage of reducing the number of operations related to the background term in the optimal estimation and of increasing their numerical stability ( $S_x$  matrix is diagonal). This representation of the state-vector in the optimal estimation is different from what is used directly by RTTOV. Therefore, the forward model  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is composed of two functions:

- the state-vector representation function  $X: \mathbb{R}^n \rightarrow \mathbb{R}^N$  and

- the RTTOV 12.1 implemented function  $f: \mathbb{R}^N \rightarrow \mathbb{R}^m$ .

Here,  $N = 304$ , corresponding to the atmospheric temperature in K, the atmospheric water vapour concentration in kg/kg, the atmospheric ozone concentration in kg/kg and the atmospheric carbon dioxide concentration in kg/kg, all at 101 fixed pressure levels each as well as the surface skin temperature. The state-vector representation function is computed individually for each component

$$[x_T \ x_W \ x_O \ x_C \ x_{T_s}] \xrightarrow{X} [X_T \ X_W \ X_O \ X_C \ X_{T_s}] \quad \text{eq(33)}$$

For temperature, we have

$$X_T = X_T^a + E_T x_T \quad \text{eq(34)}$$

where  $E_T \in \mathbb{R}^{N_T \times n_T}$  are the  $n_T$  leading principal components and  $X_T^a$  is the *a priori* temperature profile in K at 101 levels which is provided for each individual field of view.

This is similar for water-vapour (W), ozone (O) and carbon dioxide I, except that the principal components for W and O are based on  $\ln(\text{kg/kg})$  such that we get

$$X_W = \exp(\ln(X_W^a) + E_W x_W) \quad \text{eq(35)}$$

$$X_O = \exp(\ln(X_O^a) + E_O x_O) \quad \text{eq(36)}$$

$$X_C = \exp(X_C^a + E_C x_C) \quad \text{eq(37)}$$

where  $X_W^a$ ,  $X_O^a$ ,  $X_C^a$  are the *a priori* water vapour, ozone and carbon dioxide concentrations in kg/kg at the 101 fixed pressure levels.

Finally  $X_{T_s} = x_{T_s}$ .

Having defined the state-vector representation function X, we can now write the forward model function, F, used in the optimal estimation as

$$F(x) = f(X(x)) \quad \text{eq(38)}$$

and its Jacobian as

$$F'(x) = f'(X(x)) \cdot X'(x) \quad \text{eq(39)}$$

The radiances simulated by RTTOV depends on more parameters than are included as active parameters of the state-vector

- the surface pressure, which is obtained from the PWLR<sup>3</sup> retrieval
- the surface emissivity, which is obtained from the PWLR<sup>3</sup> retrieval
- the satellite zenith angle, which is obtained from the level 1 data
- the surface elevation, which is looked up from the geolocation of the pixel (§3.5.3)

### 3.8.3.2 Observations, y

IRS spectra can be represented by a small number of PC scores with only a minor loss of information. The same is true about the representation by a small number of reconstructed radiances. In fact it is easy to go from one representation to the other provided that the channel subset of reconstructed radiances is chosen such that the corresponding sub-matrix of the eigenvector matrix is non-singular. That this is always possible follows directly from the “Fundamental theorem of linear algebra”. In order to make the transformation from a subset of reconstructed radiances to PC scores numerically stable and be able to preserve all the

information available in the PC scores, it is important that the channel subset is chosen such that the condition number of this sub-matrix is kept low. In practice the condition number is heuristically minimized by choosing linearly independent rows of the eigenvector matrix with either Gaussian elimination or a Gram-Schmidt process [Strang], in which the pivot element in each step is chosen to be numerically large. This channel selection is performed separately for each of the two bands.

It is well known that the full information content in a PC score representation of a radiance spectrum can be represented as a subset of reconstructed radiances of the same size as the number of PC scores [Collard and Matricardi 2005]. Furthermore the measurement term of the optimal estimation cost-function is exactly the same whether the measurement is represented as a subset of reconstructed radiances or as PC scores – hence the retrievals performed with the two different representations are identical [Hultberg and August 2014]. The exact equivalence of the two versions of the measurement cost function term holds when the forward models used in the two cases are consistent, in the sense that the subset of radiances produced by the first model agrees with the radiances which would be obtained by expanding the PC scores obtained by the second model. Furthermore the subset of radiances must be chosen such that the observation error covariance matrix remains non-singular when transformed to the reconstructed radiance space, which, as we will see, is easily achieved.

Let there be given a truncated set of eigenvectors,  $E \in \mathbb{R}^{m \times p}$  ( $p < m$ ), based on normalised radiances

$$y = N^{-1}(y_{IRS} - y_0) \in \mathbb{R}^m \quad \text{eq(40)}$$

where

- $y_{IRS}$  is the original measurement,
- $N \in \mathbb{R}^{m \times m}$  is the noise normalisation matrix and
- $y_0$  is a reference spectrum – typically the mean of the radiance vectors found in the training set.

We proceed by writing the expressions for the measurement term of the cost functions in the two cases (the *a priori* term is not affected by the radiance representation). For this we need to introduce two forward models,  $f_{pc}: R^n \rightarrow R^p$ , taking a state vector into a PC score representation of the corresponding simulated spectrum and,  $f_{rr}: R^n \rightarrow R^p$ , taking a state vector into a subvector of the corresponding simulated spectrum, such that  $f_{rr}(x)$  is equal to the corresponding subvector of  $E f_{pc}(x)$ . The cost function for the PC score representation can now be written as:

$$(E^T y - f_{pc}(x))^T (E^T S_y E)^{-1} (E^T y - f_{pc}(x)) \quad \text{eq(41)}$$

where  $S_y$  is the observation error covariance matrix in normalised radiance space.  $S_y$  might be an identity matrix, in case the noise normalisation matrix is equal to the matrix square root of the full observation error covariance matrix (i.e. if  $N = S_y^{1/2}$ ), but most likely it is not (e.g. the normalisation account for instrument noise covariance only, or just the diagonal of it).

To express the cost function for the representation as a subset of reconstructed radiances, we let  $E_s \in R^{p \times p}$  denote the sub matrix of  $E$  obtained by selecting the rows of  $E$  corresponding to the selected channels and get

$$(E_s E^T y - f_{rr}(x))^T (E_s E^T S_y E E_s^T)^{-1} (E_s E^T y - f_{rr}(x)) \quad \text{eq(42)}$$

which, using the consistency of the two forward models ( $f_{rr}(x) = E_s f_{pc}(x)$ ), we can immediately rewrite as

$$\begin{aligned} & (E_s E^T y - E_s f_{pc}(x))^T (E_s E^T S_y E E_s^T)^{-1} (E_s E^T y - E_s f_{pc}(x)) \\ &= (E^T y - f_{pc}(x))^T E_s^T (E_s E^T S_y E E_s^T)^{-1} E_s (E^T y - f_{pc}(x)) \end{aligned} \quad \text{eq(43)}$$

If  $E_s$  is non singular, the matrix inverse can be written as a product of matrix inverses

$$\begin{aligned} & (E^T y - f_{pc}(x))^T E_s^T E_s^{-T} (E^T S_y E)^{-1} E_s^{-1} E_s (E^T y - f_{pc}(x)) \\ &= (E^T y - f_{pc}(x))^T (E^T S_y E)^{-1} (E^T y - f_{pc}(x)) \end{aligned} \quad \text{eq(44)}$$

which is identical to the PC score cost function.

The  $p$  columns of  $E$  are linearly independent (in fact they are even orthogonal) and the fundamental theorem of linear algebra tells us that a subset of  $p$  linearly independent rows exist. Furthermore it is easy to identify a linearly independent subset of  $p$  rows by Gaussian elimination or Gram-Schmidt orthogonalization.

With a channel selection obtained in this way, the cost function with both representations of the measurements (as PC scores or as a subset of reconstructed radiances) are identical, provided that the two forward models are the same, in the sense that the forward modelled PC scores expanded to radiances via pre-multiplication with the eigenvectors agree with the forward modelled radiances [Hultberg and August 2014]. We can therefore exploit the full information content in a computationally efficient way by representing the observations as a small subset of reconstructed radiances. The reconstructed radiances,  $y$ , are obtained from the IRS PC scores provided in the L1 files using the operational eigenvectors based on measurements. To suppress instrument artefacts this is followed by a projection onto the forward model subspace [Hultberg and August 2014]. For computational efficiency the matrix corresponding to the reconstruction followed by the projection is precomputed by matrix multiplication, such that only a single matrix vector multiplication is needed to obtain the double reconstructed radiances from the L1 PC scores.

Furthermore, as discussed in Appendix D, the observations must be bias corrected to get rid of systematic differences with respect to the forward model simulated radiances. The correction to apply is based on a large set of differences between observations and forward model spectra computed from collocated ECWMF analysis profiles. A simple linear model parameterized by the secant of the satellite zenith angle has been demonstrated to work well for IASI. But more advanced piecewise linear models using the PC scores as additional predictors would be better suited to take into account local biases, related to the abundances of trace gases which are assumed fixed in the retrieval, and should be explored as a possible improvement. At first, the bias corrected radiances to be ingested in the OEM,  $y_{OEM}$ , can be derived from the reconstructed radiances,  $y_{RR}$ , as follows:

$$y_{OEM} = y_{RR} + (b_0 + b_1(\zeta - \bar{\zeta}) + b_2(0mC - \overline{0mC}) + b_3(T_a - \overline{T_a})) \quad \text{eq(45)}$$

where  $b_0$ ,  $b_1$ ,  $\bar{\zeta}$ ,  $\overline{0mC}$ ,  $\overline{T_a}$  are static configuration coefficients of the linear parameterisation by the secant of the satellite zenith angle  $\zeta$ , the predicted cloud signal  $0mC$  (§3.6.2.3) and the surface air temperature  $T_a$ .

### 3.8.3.3 Observation error covariance matrix, $S_y$

The observation error covariance matrix must account for all types of errors affecting both the measurement,  $y$ , and the simulated radiances  $F(x)$ . This includes both the part of the instrument noise which is left in the reconstructed radiances and the forward model errors, for example caused by errors in the parameters which are not part of the active state-vector. As we expect higher uncertainty of the surface emissivity over land than over water, and this uncertainty affects the observation error, we use two different matrices for  $S_y$  over sea and over land.

Each of these two instances of  $S_y$  is a full matrix (i.e. with the error correlations included) computed offline as the covariance matrix of a large sample of differences between the measurement  $y$  (in the optimal estimation representation as defined above) and the corresponding forward model simulations  $F(X_{First-Guess})$  using the first-guess as input. This would seem to overestimate the observation errors, since it includes the effect of any errors in the first-guess, be it numerical forecasts or PWLR retrievals (§3.6). However in the latter configuration, since the major part of the PWLR retrieval error is smoothing error, which does not affect the radiances, the net result is just a minor overestimation. When using this “Obs minus Calc” approach to estimate the observation error covariance matrix, it is essential not to base the “Calc” on profiles which have been determined by a variational method like optimal estimation, since this would result in an underestimation of the observation errors, see [Rodgers 2000] page 190. This underestimation would be fatal, because it would occur along the directions in the range of the Jacobians, which are the directions which actually matters for the minimization. In the configuration where OEM is ran with numerical forecasts, other methods to determine this observation error covariance matrix, such as the Desroziers techniques, could be investigated closely together with the radiance assimilation community as part of Day-2 developments.

### 3.8.4 Minimization of the cost function, $J = J_x + J_y$

The total cost function to be minimised writes:

$$J = J_x + J_y = (x - x_a)^T S_x^{-1} (x - x_a) + (F(x) - y)^T S_y^{-1} (F(x) - y) \quad eq(46)$$

It is a scalar valued continuous function of  $x$  and we can apply Netwon’s method to find a zero of its derivative (the gradient). This leads to an iterative process

$$x_{n+1} = x_n - J''(x_n)^{-1} J'(x_n) \quad eq(47)$$

where  $J'(x_n)$  is the first derivative of the cost function at iteration  $n$  and  $J''(x)$  is the Hessian:

$$J''(x) = F'(x) S_y^{-1} F'(x) + S_x^{-1} \quad eq(48)$$

The iterations are stopped when the Euclidean norm of the gradient,  $\|J'(x_n)\|$ , is smaller than a small threshold value, the maximum number of iterations (5) has been reached or if the relative size of the Newton step is below  $10^{-8}$ .

$$\|J''(x_n)^{-1} J'(x_n)\| / \|x_n\| < 10^{-8} \quad eq(49)$$

In the first case we say that the process has converged. After exiting the iterations, no matter whether convergence was reached or not, the background and the observation terms of the cost function are compared to two threshold values. The current solution  $x_n$  is accepted if both cost

function terms are below its corresponding acceptance threshold, otherwise, if the optimal estimation solution is not accepted, only the first retrieval is provided.

Note that very often the Levenberg-Marquardt algorithm is used for the minimization of the cost function in implementations of optimal estimation retrievals. This leads to slower convergence (more iterations needed) than with the pure Newton method, but can converge in cases where Newton's method would fail because the initial state-vector guess,  $x_0$ , is not within the region of convergence of the method. However if we use the PWLR<sup>3</sup> retrieval for the initial state-vector guess (no matter whether PWLR<sup>3</sup> or forecast is used as *a priori*) we are already very close to the optimal estimation solution and normally well within the convergence region, such that the Levenberg-Marquardt algorithm is not needed. To be on the safe side we propose to implement a simple variant of the damped Newton method, which has been proven operationally in the IASI L2 processing. If the usual Newton step does not provide a solution with a lower value of the cost function we take a half Newton step instead

$$x_{n+1} = x_n - \frac{1}{2} J''(x_n)^{-1} J'(x_n) \quad \text{eq(50)}$$

or a quarter step, an eighth of a step, etc. until  $J(x_{n+1})$  is lower than  $J(x_n)$ .

### 3.8.5 Retrieval error covariance matrix

The retrieval error covariance matrix of the optimal estimation is given by the expression

$$(F'(x)S_y^{-1}F'(x) + S_x^{-1})^{-1} \quad \text{eq(51)}$$

evaluated at the optimal solution. We see that this is the inverse of the Hessian, which is already computed in the course of the Newton iterations. This retrieval error covariance matrix applies to the compact representation of the state vector in the optimal estimation using PC scores. Using the eigenvectors of the state vector components, it can be expanded to apply for the fixed pressure level grid. However, this expansion is better performed on the user side to save bandwidth, since the data volume is much smaller when expressing the retrieval error covariance matrix in PC space. We can save additional space by ignoring the error correlations between the state vector component, such that instead of providing an  $n$  by  $n$  matrix ( $n = n_T + n_W + n_O + n_C + 1$ ) we provide only the diagonal blocks – of dimensions  $n_T$  by  $n_T$ ,  $n_W$  by  $n_W$ ,  $n_O$  by  $n_O$  and,  $n_C$  by  $n_C$ . And finally we only need to provide the upper diagonal part of these blocks as they are symmetric.

To illustrate how the full symmetric matrix is built from the upper triangular part stored in row major order we consider a case where the number of PC scores for some parameter equals 3 and so the upper diagonal entries stored in the retrieved product are as follows:  $v_1, v_2, v_3, v_4, v_5, v_6$ .

In this case the retrieval error covariance matrix in PC space,  $S_{PC}$  would be given by this formula:

$$S_{PC} = \begin{bmatrix} v_1 & v_2 & v_3 \\ v_2 & v_4 & v_5 \\ v_3 & v_5 & v_6 \end{bmatrix} \quad \text{eq(52)}$$

For this parameter we obtain the  $N \times N$  retrieval error covariance matrix  $S_{SV}$  in pressure level space from the  $n \times n$  retrieval error covariance matrix in PC space by pre- and post-multiplication with the  $N \times n$  matrix of eigenvectors,  $E$ , as follows:

$$S_{SV} = ES_{PC}E^T \quad \text{eq(53)}$$

The unit of the retrieval error covariance matrix in pressure level spaces is K<sup>2</sup> for temperature and log(ppmv)<sup>2</sup> for water vapour and ozone.

The averaging kernels in atmospheric PC space,  $A_{PC}$ , can be computed with this formula:

$$A_{PC} = I - S_{PC}S_x^{-1} \quad \text{eq( 54)}$$

where  $S_x$  is the static background error covariance matrix  $S_x$  used in the cost function.

From this, the averaging kernels in the pressure grid space can be derived as follows:

$$A_{SV} = EA_{PC}E^T \quad \text{eq( 55)}$$

We note that this retrieval error estimate ignores the fact that the *a priori* could be depending on the measurement if the first retrieval is used instead of ECMWF forecasts. This can be taken into account by combining the averaging kernels of the first and second retrievals as described in detail in [Hultberg and August 2016]. However, since this improved error estimate is not yet operational with IASI, we do not specify it for Day-1 of MTG-IRS.

### **3.9 Quality control**

All parameters resulting from first (PWLR<sup>3</sup>, section 3.6) and second (OEM, section 3.8) retrievals shall be verified against configurable validity bounds. In addition, the temperature and humidity profiles from these two methods which would exceed the super-saturation and super-adiabatic conditions shall be modified and flagged accordingly to reflect that unphysical conditions were met, as described in the sections 4.10 and 4.12.

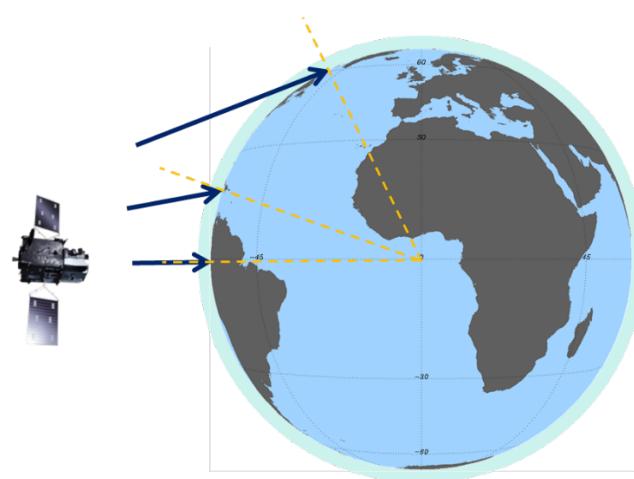
### **3.10 Post-processing**

#### **3.10.1 Geometric corrections**

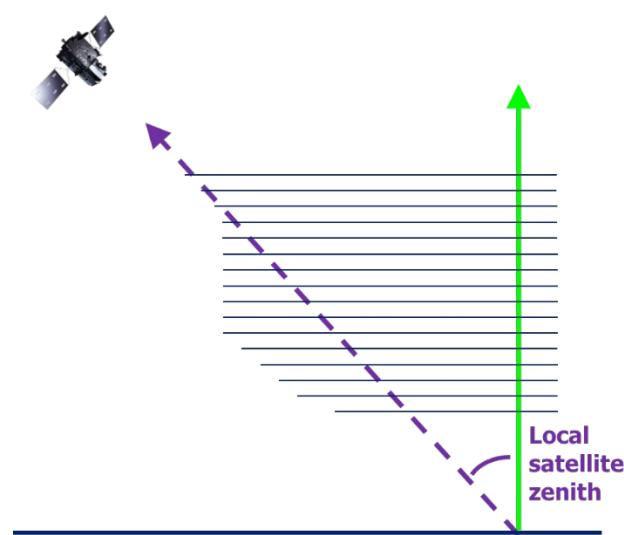
##### **3.10.1.1 Reconstruction of vertical profiles**

While the profiles delivered to the Users are intended at the vertical of the target point (green profile in the Figure 27), the geophysical parameters are effectively retrieved along the line of sight (in violet in Figure 27) with the approach described in the present version of this document. This may require staging retrievals to combine information from adjacent dwells and is an extra layer of complexity in the processing. Alternative algorithms aiming direct retrieval of vertical profiles (with statistical methods or 3D-variational methods) will be investigated in future to circumvent this.

The parallax effect indeed becomes important as the observation target departs from the sub-satellite point. For instance, observations in London, in the Netherlands or centre of Germany correspond to satellite zenith angle of about 60°, in which case the parallax effect for levels at 5 and 10 km correspond to approximately 1 and 2 pixels shift, respectively. Similarly, the ozone information retrieved at 30 km would be reported about 6 pixels away from its true location if the parallax effect is not taken into account.

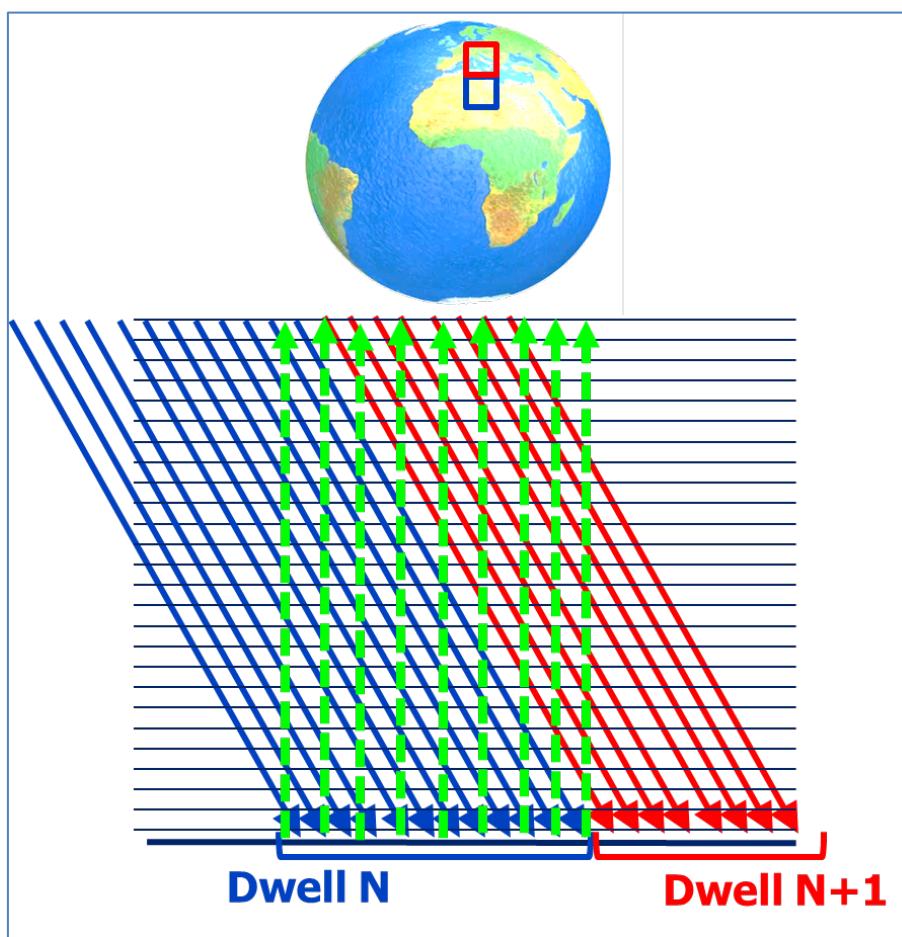


*Figure 26: Schematic view of IRS satellite zenith angle wrt local zenith (yellow).*



*Figure 27: Observed slanted (purple) and vertical profile provided to the End-Users (green).*

The purpose of this post-processing step is to reconstruct the information at the vertical of the target point (i.e. the pixel centre) from the slanted retrievals which are intercepting it. The Figure 28 illustrates the principle of the reconstruction, which may also require combining information from adjacent dwells.



*Figure 28: Reconstructing profiles at the vertical of a pixel with slant retrievals.*

The same pressure grid is used for the end product (vertical profile, green in Figure 28) and the intermediate output slanted retrievals from PWLR<sup>3</sup> or OEM (blue and red in Figure 28). While the collocation slanted-to-vertical profiles can be systematically computed online after §4.16, it may represent unnecessary large computation overhead.

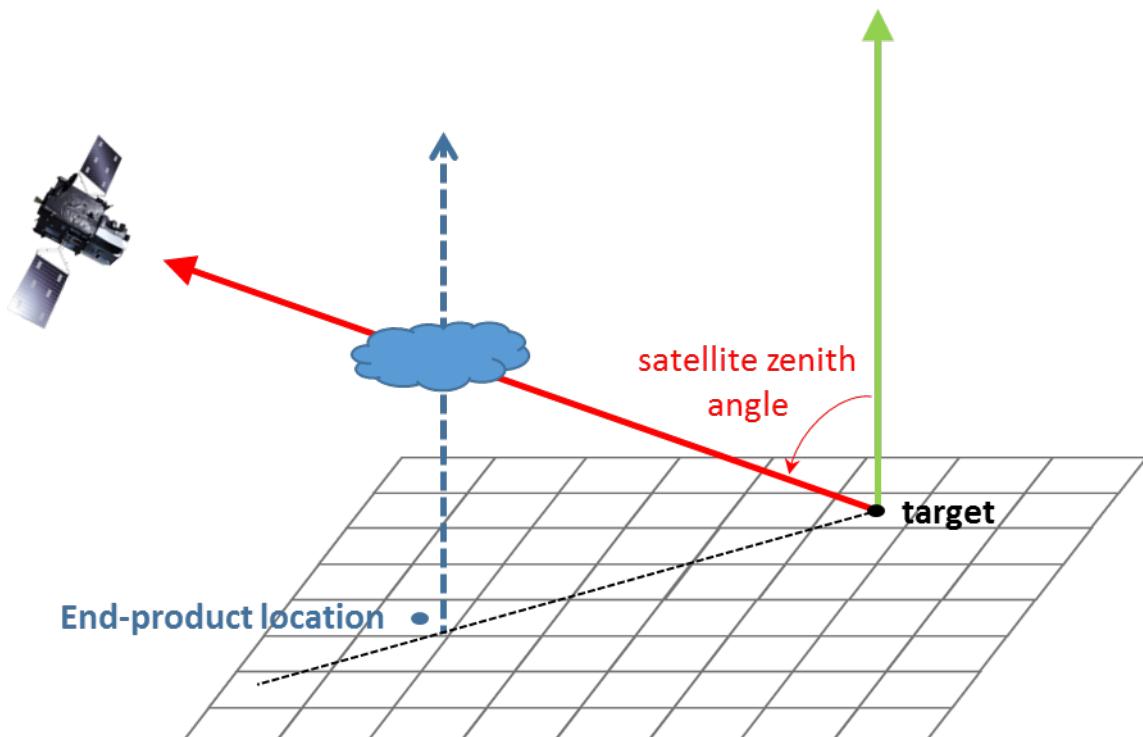
The correspondence slanted-to-vertical within any given dwell can be assumed to follow a deterministic pattern, which is function of the viewing geometry for that dwell. The assignment of atmospheric profiles parameters at given vertical pressure levels is hence alternatively approximated via pre-computed look-up tables (LUT), by off-line evaluation of the profiles levels' geolocation (and hence nearest IRS grid point) as explained in §4.16. Slanted retrievals from adjacent dwells may also be required to compose vertical profiles. In such cases, a fixed nominal relative position is assumed at first order and slanted-to-vertical correspondence table across dwells can be pre-computed off-line.

For each pixel of a given dwell and for each pressure level of the end-product vertical profiles, an auxiliary dataset contains the dwell identifier and pixel indices of the associated slanted retrieval. In case of recombination from adjacent dwells, the actual positions of anchor points (e.g. dwell corners...) is evaluated relatively to the nominal collocation and offsets in pixels in the target detector matrix is computed. If significant, the offsets are applied to the pre-computed look-up tables to perform construction of vertical profiles across dwells.

Overall, different slanted-to-vertical LUTs may need to be pre-configured depending on the satellite position (essentially elevation) around its nominal orbital position.

### 3.10.1.2 Geolocation of the clouds

The Figure 29 illustrates the effect of the viewing geometry on the perceived and actual location of a cloud retrieved on a slant path. The Earth coordinates of the sub-point of a cloud retrieved along a slant path can be computed following the methodology explained in section 4.16, using the pressure to height conversion in §4.3. The cloud can then be assigned to the nearest IRS pixel of the same dwell and/or in a neighbouring dwell as applicable.



*Figure 29: Apparent and actual geolocation of a cloud retrieved on a slant path (IRS pixels on ground illustrated with the black boxes)*

### 3.10.1.2.1 Computation of columnar amounts

The total columns of water-vapour and ozone are computed as described in section 4.2, using the profiles retrieved with the statistical (§3.6) and optimal estimation (§3.8) methods, after application of the geometrical corrections (§3.10.1.1) where necessary. These total columns are intended in the final products together with the atmospheric profiles, as presented in Appendix C.

2D fields of partial columnar amounts of water-vapour may also be pre-computed from the profiles to serve users interest. The principle was addressed in the ESSL Testbed 2018 ([www.essl.org](http://www.essl.org)) where NWC SAF proposed low and mid-tropospheric moisture integrated content as new products derived from SEVIRI. Similarly, layer amounts can be produced from IRS L2 profiles and are defined in the following section.

### 3.10.1.2.2 Computation of stability indices

Convection can yield severe weather events (thunder, wind gust, tornadoes, hail, flash-flood...) which can have dramatic societal consequences. Identifying areas with potential atmospheric instabilities is hence critical to issue accurate warnings, as early as possible, to prepare population, economic actors and civil protection. As identified in [Eyre et al. 2011], 2D fields of stability indices are important outputs expected from IRS sounding for forecasting purposes.

Following the review of the IRS L2 product content proposal, it was confirmed in the 5<sup>th</sup> MAG meeting that stability indices are to be generated centrally by the IRS L2 processor. The possibility to retrieve the stability indices directly from the observations with a statistical method, here the PWLR<sup>3</sup> (section 3.6), and their potential will be investigated. It is meanwhile proposed to derive these indicators in the classical way from the retrieved profiles, after their geometrical post-processing (§3.10.1.1).

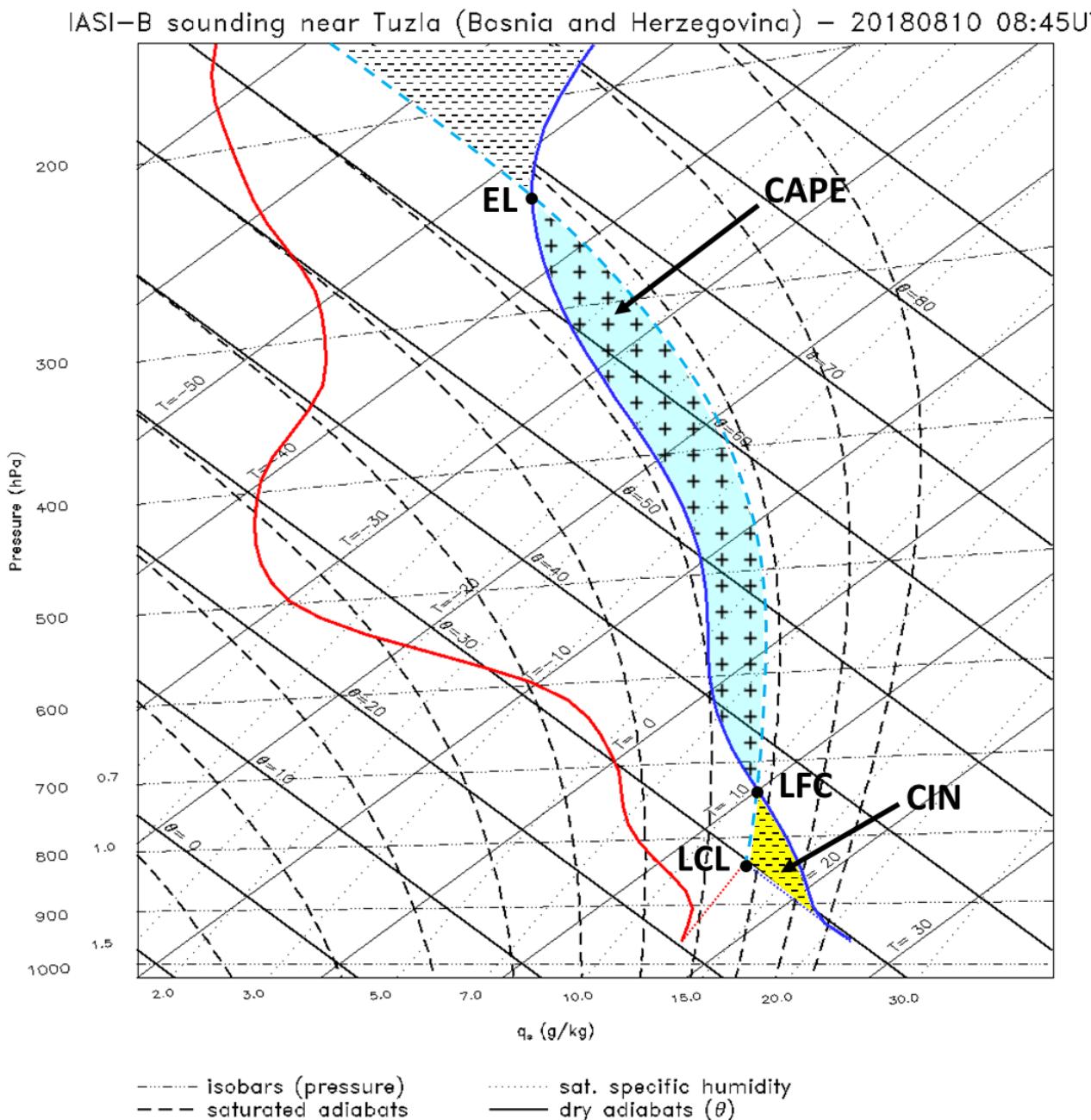
The OEM retrievals should be considered in priority and PWLR<sup>3</sup> profiles be used when the former information is not available.

There exists a large diversity of stability indicators which have been conceived from atmospheric sounding profiles to assess thunderstorm risks, historically from *in situ* radiosonde, e.g. [Haklander and Delden 2003]. Their significance to predicting storms have been assessed through their correlation with actual records of severe weather events (wind gusts, lightning, hail, flood...) in difference contexts, see e.g. [Haklander and Delden 2003][Kunz 2007][Putsay et al. 2017]. Storm forecasting however practically involves complex analyses of several numerical model outputs, completed with ground- and space-based observations, to evaluate the dynamic evolution of weather systems through a range thermodynamic parameters. This includes snapshots anterior to the time the forecasts are being made and projections into the future. The wind information is also an important ingredient of now- and short-term forecasting. It is hence difficult in such studies to try to relate directly actual severe weather developments to a particular stability index taken in isolation or regarded as a static representation. Also, different combination of kinetic parameters may be regarded depending on the location and on the weather situations. The Table 5 therefore proposes an initial set of stability indices to be generated, which can be reviewed in future as more experience will be made with e.g. Polar satellite sounding products. This list aims continuity with MSG GII heritage and consistency with the MTG-FCI follow-up products [FCI L2 PS]. It is completed with a few more indices collected in the literature and from initial interactions with forecasters.

Identifier	Full name	Short description
LI	Lifted-index [Galway 1956]	Difference between the environment temperature at 500 and the temperature of a surface-air parcel lifted adiabatically to that level.
K-index	K-index [George 1960]	Combined index assessing the potential for thunderstorm development concerning mid-level temperature lapse rate (between 500 and 850 hPa levels) and humidity (850 hPa dewpoint, dewpoint depression, DD at 700 hPa).
LPW	Layer Precipitable Water	Partial columnar amounts of water-vapour in layers. The boundaries are configurable and initially defined as [surface to 850 hPa] ; [850 to 500 hPa] ; [500 hPa to top of atmosphere]
Dθe	DTHETA <sub>E</sub> [Atkins and Wakimoto 1991]	Used to diagnose areas with vertically decreasing equivalent potential temperature, which are considered to be conditionally unstable.
MB	Maximum Buoyancy [König et al. 2001]	Similar to DTHETA <sub>E</sub> , looking for vertically decreasing equivalent potential temperature in a larger vertical domain.
SBCAPE or CAPE	Surface-Based Convective Available Potential Energy	Amount of potential energy available for convection to an air parcel theoretically lifted to the level of free convection (LFC) and which would further ascent from its own buoyancy. It also relates to the maximum potential vertical speed within an updraft. Small or negative CAPE values are indicative of stable atmospheres.
MLCAPE	Mixed-Layer CAPE	Same concept as SBCAPE, but evaluated with the air parcel average located in the lowest 100-mb. It is commonly used to assess instability when the atmosphere is well mixed (e.g. in the afternoon)
MUCAPE	Maximum Unstable CAPE	It is the maximum of CAPE values computed for every level in the first 300 to 500 hPa (upper limit configurable). It helps assessing the possibility of elevated convective cells in case of low level inversions (e.g. at night or behind a cold front).

*Table 5: List of stability indices in IRS L2 products*

The corresponding algorithms for their computation are detailed in the following subsections.



**Figure 30: Conditional instability illustrated with a real atmospheric sounding from IASI-B near Tuzla (Bosnia and Herzegovina) on 10 August 2018.**

We present hereafter and with the Figure 30 a simple overview of the ingredients involved in the computation of stability indices and assessment of convective mechanisms. It is illustrated with a real atmospheric sounding from IASI-B on 10/08/2018 near Tuzla (Boasnja and Herzegovina), where the retrieved temperature profiles is drawn in blue and humidity in red. The Lifting condensation level (LCL) is the level where the surface air parcel theoretically lifted in an adiabatic way reaches saturation (along the dry adiabat, solid line top-left to bottom right diagonal on Figure 30). Then, if theoretically further lifted, this air parcel would follow the saturated adiabats on the tephigram until it reaches the Level of Free Convection (LFC). From there onwards, the air parcel –warmer therefore less dense than the environment- would

continue rising from its own buoyancy until the temperature of the rising parcel meets the environment temperature and hence the Equilibrium Level (EL). The acceleration that air parcel will experience can be related to the difference in temperature with respect to the temperature of the surrounding air at each successive level. We hence define the convective available potential energy (CAPE) as the surface between the rising parcel temperature (dash cyan line) and that of the environment (blue profile) between LFC and EL. It is displayed with the ‘plus’ sign on cyan background. On the contrary, the area delimited below the LFC corresponds to the amount of energy necessary to force the surface air parcel to rise (as it remains cooler and so denser than the environment). It is the Convective Inhibition (CIN) and is represented with the ‘minus’ sign on yellow background.

### 3.10.1.3 Lifted Index

The Lifted-index  $LI$  expresses the difference between the theoretical temperature of an air parcel lifted at 500 hPa and its environment. It is computed as follows

$$LI = T_{air} - T_{air}^{lifted} \quad \text{eq(56)}$$

Where

- $T_{air}$  is the air temperature at 500 hPa, interpolated logarithmically with pressure from the retrieved profiles between the adjacent pressure levels as described in §4.1;
- $T_{air}^{lifted}$  is the theoretical temperature of the air parcel lifted from surface to 500 hPa.

The surface air parcel is defined as the average (linearly with pressure) temperature and humidity content of the lowest 100 hPa in the atmosphere, with the following notation:

- $T_{sfc}$  the surface air temperature
- $q_{sfc}$  the surface air specific humidity
- $p_{sfc}$  the surface air pressure

The surface air parcel theoretically lifted almost always reaches the condensation level (LCL) along the dry adiabat before the 500 hPa level and then follows the moist adiabat (see Figure 30). The computation of the temperature and pressure at LCL,  $T_{LCL}$  and  $p_{LCL}$ , is described in §4.15.

In the rare events where  $p_{LCL}$  is lower than 500 hPa (i.e. LCL is higher than 500 hPa in altitude), the temperature of the near-surface air parcel simply follows the dry adiabat until 500 hPa. The lifted air temperature  $T_{air}^{lifted}$  in eq( 56) can then be inferred directly from eq( 109) in §4.13, with  $T_{sfc}$ ,  $p_{sfc}$  and  $p_{air}^{lifted} = 500 \text{ hPa}$ .

In most situations, the lifting condensation level is well below the 500 hPa height (i.e.  $p_{LCL} > 500 \text{ hPa}$ ) and the air parcel follows the moist adiabat (dash line in Figure 30) between  $p_{LCL}$  and 500 hPa. Firstly, the equivalent potential temperature for the surface air parcel,  $\theta_e^{sfc}$ , is calculated using  $T_{LCL}$  and the humidity at surface following eq( 111) in §4.13.  $T_{air}^{lifted}$  in eq( 56) is then computed as the difference:

$$T_{air}^{lifted} = T_1 - T_2 \quad \text{eq(57)}$$

$T_1$  is the temperature that a totally dry air parcel of temperature  $\theta_e^{sfc}$  at 1000 hPa would have if adiabatically displaced to 500 hPa. It is calculated after *eq( 109)* with 1000 and 500 hPa as pressure limits,  $\theta_e^{sfc}$  as origin temperature and assuming totally dry air, i.e.  $q=0$  in *eq( 106)* and *eq( 107)*;

$T_2$  is computed using an empirical regression scheme [Doswel *et al* 1982]. The following steps are involved:

### 3.10.1.3.1 Compute t

$$t = T_1 - 293.16 \quad \text{eq( 58)}$$

### 3.10.1.3.2 If $t < 0$ then :

$$T_2 = \frac{15.13}{P^4} \quad \text{eq( 59)}$$

### 3.10.1.3.3 else:

$$T_2 = \frac{29.93}{P^4} + 0.96t - 14.8 \quad \text{eq( 60)}$$

In both equations *eq( 59)* and *eq( 60)*, the term P is a third order polynomial of t:

$$P = 1 + c_1 t + c_2 t^2 + c_3 t^3 \quad \text{eq( 61)}$$

	$t \leq 0$	$t > 0$
c <sub>1</sub>	-8.8416605 E-03	+3.6182989 E-03
c <sub>2</sub>	+1.4714143 E-04	-1.3603273 E-05
c <sub>3</sub>	-9.6719890 E-07	+4.9618922 E-07

*Table 6: Values of c<sub>1</sub>, c<sub>2</sub>, c<sub>3</sub> (dependent on the value of t)*

This final  $T_{air}^{lifted}$  of *eq( 57)*, is used in equation *eq( 56)* to complete the computation of the Lifted Index.

Negative *LI* values correspond to a surface air parcel potentially warmer than the environment at 500 hPa if lifted to this level, and therefore unstable.

In case the surface pressure is below 500 hPa, the lifted index is not defined.

### 3.10.1.4 K-Index, KI

The K-index,  $KI$ , is a combined index to assess the potential for thunderstorm development based on mid-level temperature lapse rate and humidity [George 1960]. It writes:

$$KI = (T_{air}^{850} - T_{air}^{500}) + TD_{air}^{850} - (T_{air}^{700} - TD_{air}^{700}) - 273.15 \quad \text{eq(62)}$$

Where

- $T_{air}^p$  is the atmospheric temperature at pressure level  $p = \{850, 700, 500\}$  hPa. The temperature at the adjacent respective pressure levels in the retrieved profile are interpolated to the target pressure, logarithmically with pressure (§4.1).
- $TD_{air}^p$  is the dew point temperatures at  $p = \{850, 700\}$  hP. The humidity at the adjacent respective pressure levels in the retrieved profile are interpolated to the target pressure, logarithmically with pressure (§4.1), and converted to dew point temperature as described in section 4.9.

All temperature values in **eq( 62)** are expressed in Kelvin, the K-Index, however, is commonly expressed in deg Celsius, which explains the term 273.15 subtracted in that equation.

In case the surface pressure is below 850 hPa, the K-Index is not defined.

### 3.10.1.5 Layer-Precipitable Water, LPW

The LPW shall contain the vertically integrated humidity partial columns in a limited number of atmospheric layers. The actual number of layers and their boundaries shall be configurable. As an initial configuration, the following three layers are defined:

- i. Surface to 850 hPa
- ii. 850 to 500 hPa
- iii. 500 hPa to top of atmosphere

The integration of the partial columns is performed with the retrieved profiles as described in section 4.2.

### 3.10.1.6 Maximum Buoyancy, MB

After [König et al. 2001], the maximum buoyancy index,  $MB$ , aims the identification of areas with vertically decreasing equivalent potential temperature, which are considered to be conditionally unstable. It is the difference between the maximum equivalent potential temperature in the lower troposphere, from surface to 850 hPa, and the minimum equivalent temperature in the free troposphere from 700 to 300 hPa. It writes:

$$MB = \max \theta_e]_{SFC}^{850 \text{ hPa}} - \min \theta_e]_{700 \text{ hPa}}^{300 \text{ hPa}} \quad \text{eq(63)}$$

The equivalent potential temperature  $\theta_e$  are computed from the retrieved profiles as described in **eq( 111)** and section 4.13.

If the surface pressure is below 850 hPa, the MB index is not computed.

### 3.10.1.7 DTHETA<sub>e</sub>, Δθ<sub>e</sub>

The DTHETA<sub>e</sub> and maximum buoyancy (§3.10.3.4) indexes are similar diagnostic parameters, diagnose conditional instability with the maximum gradient of equivalent potential temperature between surface and elevated layers. DTHETA<sub>e</sub> was first defined [Atkins and Wakimoto 1991] and declined in several versions with different layer configurations where the maximum/minimum equivalent potential temperature are looked for. Dry air decreases the equivalent potential temperature, hence DTHETA<sub>e</sub> can be high not only due to steep temperature lapse rate but also because of presence of dry layers in the mid-troposphere. Thus, high DTHETA<sub>e</sub> can be sometimes favourable also for convection with strong downdrafts. In this approach, DTHETA<sub>e</sub> is evaluated as follows:

$$\Delta\theta_e = \max\theta_e|_{SFC} - \min\theta_e|_{SFC}^{500\text{ hPa}} \quad \text{eq(64)}$$

Where  $\max\theta_e|_{SFC}$  is the maximum equivalent potential temperature  $\theta_e$  evaluated according to **eq( 111)** and section 4.13 in the first 100 hPa above surface and  $\min\theta_e|_{SFC}^{500\text{ hPa}}$  is the minimum equivalent potential temperature evaluated between surface and 500 hPa.

If the surface pressure is below 500 hPa, the DTHETA<sub>e</sub> parameter is not defined.

### 3.10.1.8 Available Potential Energy indices, CAPE and CIN

The acceleration  $\vec{a}$  experienced by an air parcel due to density differences (buoyancy acceleration) with surrounding air can be related to the differences in temperature between the air parcel,  $T_{parcel}$ , and in its environment,  $T_{env}$ .

$$\vec{a} = \frac{(T_{parcel} - T_{env})}{T_{env}} \vec{g} \quad \text{eq(65)}$$

$\vec{g}$  being the gravitational field.

When the parcel is warmer than its environment (less dense), the buoyancy is positive and in the absence of other processes, the air parcel would naturally rise. The integral of the theoretical accelerations an air parcel would experience if theoretically moved between two atmospheric levels relates to the available potential energy (APE) [Lorenz 1955]. It can be written between the altitudes  $z_1 < z_2$  or the atmospheric pressure levels  $p_1 > p_2$  as:

$$APE = \int_{z_1}^{z_2} \frac{(T_{parcel} - T_{env})}{T_{env}} g dz = -R_d \int_{p_1}^{p_2} (T_v^{parcel} - T_v^{env}) d \ln p \quad \text{eq(66)}$$

Where  $T_v$  is the virtual temperature (§4.10).

#### CIN:

When the integral is negative, an external source of energy is necessary to move the parcel upwards. These negative areas inhibit the convection and we define the Convective INhibition as the area between the parcel origin and the LFC (yellow area in Figure 30), if there is one.

#### CAPE:

Conversely, positive areas correspond to convective available potential energy for the parcel (positive buoyancy). We define here the convective available potential energy (CAPE) as the integral between the LFC and EL (cyan area in Figure 30).

The evaluation of the differences between environment and parcel temperatures along the theoretical ascent and the computation of CIN and CAPE are described hereafter.

### 3.10.1.8.1 Procedure to computing convective inhibition and available potential energy

For a given retrieved profile of temperature  $T$  and water-vapour mixing ratio  $r$  (see  $r_{H_2O}$  in §4.9) defined on a pressure grid with  $N_{lev}$  levels, let  $i$  be the origin level of the air parcel for which the available potential energy will be assessed, with the corresponding pressure  $p_i$ , temperature  $T_i$  and humidity  $r_i$ :

I. Characterise the LCL (§4.15) corresponding to the air parcel at original level  $i$

II. Compute the virtual temperatures  $T_{v,i}^{env} = T_{v,i}^{parcel}$  at level origin  $i$ , with  $T_i$  and  $r_i$  in eq( 105) of §4.10

III. Let initialise the following variables

CIN=0 and CAPE=0

A temporary variable sum  $\zeta = 0.0$

$r_0^G = r_i$

$T_0^G = T_i$

$s_{p,i}$ , the pseudo adiabatic entropy at pressure level  $i$  [Bryan 2008]

$$s_{p,i} = 1005.7 * \ln T_i - 287.04 * \ln(p_i - e_{v,i}) + \frac{2.501 * 10^6 - 2320(T_i - 273.15)}{T_i} * r_i \quad \text{eq( 67 )}$$

$$- 461.5 * r_i \ln(e_{v,i}/e_{sat,i})$$

With  $e_{sat,i}$  the water-vapour saturation pressure as explained in section 4.9 eq( 93 ) and  $e_{v,i}$  is the partial pressure of water-vapour at level I, computed after eq( 99 ) in section 4.9.

IV. LOOP upward above level  $i$  through each level  $j$  ( $p_j < p_i$ ) of the retrieved profile (up to 100 hPa)

i. Compute the environment virtual temperature  $T_{v,j}^{env}$  with  $T_j$  and  $r_j$  in eq( 105) of §4.10

ii. Compute the theoretical virtual temperature  $T_{v,j}^{parcel}$  for the air parcel

**IF**  $p_i \geq p_{LCL}$  **THEN** the parcel follows the dry adiabat

- a. Compute  $T_{lifted,j}^{parcel}$  the temperature of the parcel lifted at pressure  $p_j$ , using **eq( 109 )** of §4.13 with temperature  $T_i$  and pressure  $p_i$  at origin level  $i$ .
- b.  $T_{v,j}^{parcel}$  is then computed with  $T_{lifted,j}^{parcel}$  and  $r_i$  in eq( 105) of §4.10

**ELSE** the lifted parcel follows the moist adiabat. The lifted parcel temperature and mixing ratio along the pseudo adiabatic ascent up to level  $j$  is solved iteratively.

- Let compute  $r_j^{sat}$ , the saturation mixing ratio at level  $j$ . First evaluate the water-vapour saturation pressure  $e_{sat,j}$  at level  $j$  with environment temperature  $T_j$  after eq( 93 ) of section 4.9. Then derive  $r_j^{sat}$  from using  $e_{sat,j}$ .
- Let compute  $L_j^{vap} = 2.501 * 10^6 - 2320(T_j - 273.15)$ , **eq( 68 )** the latent heat of vaporisation in the environment at pressure level  $j$ .
- Let compute  $s_{p,j}^L$ , an intermediate factor of the pseudo-adiabatic entropy

$$s_{p,j}^L = \left( 1005.7 + 4190 * r_j^{sat} + \frac{L_j^{vap} * L_j^{vap} * r_j^{sat}}{461.5 * T_j * T_j} \right) / T_j, \quad \text{eq( 69 )}$$

- Let initialise

The iterated lifted temperature  $T_j^G = T_j$

The iterated lifted mixing ratio  $r_j^G = r_j^{sat}$

- Solve iteratively the conditions for the lifted parcel on the moist adiabat

**LOOP** on configurable number of iterations (*typically 7 maximum for convergence*)

$$\cdot \quad \text{Compute } c_{pw} = \zeta + 4190 * \frac{1}{2}(r_j^G + r_0^G) * \ln \frac{T_j^G}{T_0^G} \quad \text{eq( 70 )}$$

- . Compute  $e_{v,j}^G$  the iterated water-vapour partial pressure using  $r_j^G$  and  $p_j$  in eq( 93 ) of section 4.9.

- . Compute the pseudo-adiabatic entropy increment  $s_{p,j}^G$  as

$$s_{p,j}^G = 1005.7 * \ln T_j^G - 287.04 * \ln(p_j - e_{v,j}^G) + \frac{2.501*10^6 - 2320(T_j^G - 273.15)}{T_j^G} r_j^G + c_{pw} \quad \text{eq( 71 )}$$

- . Update  $T_j^G$ , the lifted temperature  $T_j^G += (s_{p,i} - s_{p,j}^G)/s_{p,j}^L$   $\text{eq( 72 )}$

- . Compute  $e_{sat,j}^G$  the saturated water-vapour pressure in the lifted parcel, using updated  $T_j^G$  in eq( 93 ) of section 4.9.

- . Update  $r_j^G$ , the lifted mixing ratio using  $e_{sat,j}^G$  in of section 4.9.

**ENDLOOP**

Set  $T_{lifted,j}^{parcel} = T_j^G$  and  $r_{lifted,j}^{parcel} = r_j^G$

$T_{v,j}^{parcel}$  is then computed with  $T_{lifted,j}^{parcel}$  and  $r_{lifted,j}^{parcel}$  in eq( 105 ) of §4.10

Set  $r_0^G = r_j^G$ ,  $T_0^G = T_j^G$  and  $\zeta = c_{pw}$  for the next pressure level.

**ENDELSE**

- iii. Compute the available potential energy for the layer defined by  $p_{j-1}$  and  $p_j$

$$APE_j = R_L \left( \frac{(T_{v,j-1}^{env} + T_{v,j}^{env})}{2} - \frac{(T_{v,j-1}^{parcel} + T_{v,j}^{parcel})}{2} \right) * \frac{(p_{j-1} - p_j)}{(p_{j-1} + p_j)} * 2 \quad \text{eq( 73 )}$$

where  $R_L = 287.06 \text{ J}. \text{kg}^{-1}. \text{K}^{-1}$  is the specific gas constant for dry air.

- iv. Test  $T_{lifted,j}^{parcel}$  vs  $T_j^{env}$  to identify a potential level of free convection and then equilibrium level.

WHILE  $(T_{lifted,j}^{parcel} < T_j^{env})$  -i.e. the level  $j$  is below LFC-:

$$CIN += APE_j \quad \text{eq( 74 )}$$

Then WHILE  $(T_{lifted,j}^{parcel} > T_j^{env})$  -i.e. the level  $j$  is above LFC and below EL-  
 $CAPE += APE_j$   $\text{eq( 75 )}$

The upward iterations on level  $j$  must stop once the EL is met or if  $p_j < 100 \text{ hPa}$ .

**ENDLOOP** on level  $j$

### 3.10.1.8.2 Surface-based CAPE (SBCAPE or CAPE) and CIN

The surface-based CAPE, also referred to as SBCAPE or CAPE for simplicity here, and CIN are computed following the procedure described in §3.10.3.6.1, where the air parcel at origin is defined (temperature, humidity, pressure) in the first atmospheric level above surface.

### 3.10.1.8.3 Mixed-Layer CAPE, MLCAPE

The mixed-layer CAPE, or MLCAPE, is computed following the procedure described in §3.10.3.6.1, where the air parcel at origin is defined as the average (linearly with pressure) temperature, humidity and pressure in the first 100 hPa above surface.

### 3.10.1.8.4 Maximum Unstable CAPE, MUCAPE

The maximum unstable CAPE, or MUCAPE, is the maximum of the individual CAPE values computed following the procedure described in §3.10.3.6.1 for every level in the first 300 .



## 4 GENERIC FUNCTIONS

This section describes the algorithms that may be used at different places of the IRS L2 processing sequence.

### 4.1 Vertical interpolations of atmospheric temperature and constituent profiles

The temperature, water-vapour and ozone profiles are interpolated according to the following interpolation schemes in the standard 101-pressure grid.

The temperature varies approximately linearly with the height, whereas the pressure decreases exponentially with height, so that a suitable interpolation of temperature  $T$  at pressure  $p$  in a pressure-level grid reads as follows in *eq( 76)*:

$$T = T_0 + \frac{T_1 - T_0}{\ln\left(\frac{p_1}{p_0}\right)} \ln\left(\frac{p}{p_0}\right) \quad \text{eq( 76)}$$

where  $p_0$  and  $p_1$  are pressure levels surrounding  $p$ , and  $T_0$  and  $T_1$  are the corresponding temperature values. The CO<sub>2</sub> profiles is interpolated in the same way as temperature but the grid can be different.

The water-vapour mixing ratio  $W$  decreases exponentially with height, so that the interpolation at  $p$  on a pressure grid is double-logarithmic:

$$\ln(W) = \ln(W_0) + \frac{\ln\left(\frac{W_1}{W_0}\right)}{\ln\left(\frac{p_1}{p_0}\right)} \ln\left(\frac{p}{p_0}\right) \quad \text{eq( 77)}$$

where  $p_0$  and  $p_1$  are pressure levels surrounding  $p$ , and  $W_0$  and  $W_1$  are the corresponding water-vapour values. Ozone profiles are interpolated in the same way as water-vapour.

### 4.2 Vertical integration of atmospheric constituent concentrations

To evaluate the total column TC of an atmospheric constituent, two quantities have to be evaluated:

- Partial columnar amount  $c_i$  between two pressure levels  $p_i > p_{i+1}$
- Partial columnar amount  $c_{surf}$  between the surface pressure and the first pressure level  $p_1$

Let  $p_i$  and  $\mu_i$  be the pressure and mass mixing ratio of a given atmospheric constituent defining a vertical profile on  $N_{lev}$  levels. With  $p_i$  and  $\mu_i$  expressed in hPa and kg/kg, respectively, the partial columnar amount  $c_i$  in kg/m<sup>2</sup> of this atmospheric constituent in the layer between two pressure levels  $p_i > p_{i+1}$  is given by the formula

$$c_i = \frac{\bar{\mu}(p_i - p_{i+1}) * 100}{g(\bar{z}, \phi)}; i = 0..N_{lev} - 2 \quad \text{eq( 78)}$$

Where

$$\bar{\mu} = (\mu_i + \mu_{i+1})/2 \quad \text{eq( 79)}$$

$$\bar{z} = (z_i + z_{i+1})/2 \quad \text{eq( 80)}$$

Are the average mixing ratio and altitude of the layer defined by the pressure levels  $p_i$  and  $p_{i+1}$ .

$g(\bar{z}, \phi)$  is the local gravitational acceleration depending on geodetic latitude,  $\phi$ , and height,  $\bar{z}$ , as defined in 4.3, eq(87). The conversion from pressure levels to height is done as described in 4.3, starting from the surface level, whose elevation is known (3.5.3). Alternatively, it may also be approximated as follows:

$$\bar{z} = -4000(\log\left(\frac{p_i}{1013}\right) + \log\left(\frac{p_{i+1}}{1013}\right)) \quad \text{eq( 81)}$$

To compute the integrated total columnar amount  $TC$  of an atmospheric constituent, all partial columnar amounts of layers between the successive pressure levels on which the mixing ratio profile is provided must be computed as indicated and summed.

To compute the partial columnar amount  $c_{surf}$  in the layer between the surface pressure and the first pressure grid level above the surface on which the profile is defined, the mixing ratio is assumed constant in this layer and equal to the mass mixing ratio at the first pressure grid level above the surface. Hence:

$$c_{surf} = \mu_0(P_S - p_0)/g(\bar{z}, \phi) \quad \text{eq( 82)}$$

Where  $P_S$  is the surface pressure and  $\bar{z} = (z_{surf} - z_0)$ , with  $z_{surf}$  the surface elevation and  $z_0$  the height of the first pressure grid level.

Finally, the total column writes:

$$TC = c_{surf} + \sum_{i=0}^{N_{lev}-2} c_i \quad \text{eq( 83)}$$

#### 4.3 Conversion from pressure to height levels

The conversion between pressure and height levels is done with the barometric equation (hydrostatic equilibrium). It is important that this equation is applied layer-wise as it represents an integration where constant temperature and humidity is assumed between the two consecutive levels defining a layer. The height difference between two levels  $i$  and  $j$  with known pressures  $p_i$  and  $p_j$  is given by the following:

$$z_i - z_j = \frac{-R_L \bar{T}_v}{g(z_i, \phi)} \cdot \ln\left(\frac{p_i}{p_j}\right) \quad \text{eq( 84)}$$

Alternately, for known heights and the ratio of the pressures at levels  $i$  and  $j$  is

$$\frac{p_i}{p_j} = \exp\left(\frac{-g(z_i, \phi)(z_i - z_j)}{R_L \bar{T}_v}\right) \quad \text{eq( 85)}$$

where

$R_L = 287.06 \text{ J. K}^{-1} \cdot \text{kg}^{-1}$  is the gas constant for dry air,  
 $\bar{T}_v$  is the layer mean virtual temperature, represented as

$$\bar{T}_v = (T_{v,i} + T_{v,j})/2 \quad \text{eq( 86)}$$

Where  $T_{v,i}$  is computed as described in 4.10 with eq( 105)

The acceleration due to gravity  $g$  (in m.s<sup>-2</sup>) is a function of geographic latitude  $\phi$  and height  $z$  (in m):

$$g(z, \phi) = 9.80616(1 - 0.0026373 \cos(\phi) + 0.0000059 \cos^2(\phi)) \\ - (3.085462 * 10^{-6} + 2.27 * 10^{-9} * \cos(\phi))z \\ + (7.254 * 10^{-13} + 10^{-20} * \cos(\phi))z^2 \\ - (1.517 * 10^{-19} + 6 * 10^{-22} * \cos(\phi))z^3 \quad \text{eq( 87)}$$

The form of eq( 84) and eq( 85) requires that the calculations always begin at a level with known start values of  $z$  and  $p$ . Usually, surface pressure and surface height are the known lower boundary values and Equations eq( 84) and eq( 85) are iterated towards the top of the atmosphere. The surface height for each MTG-IRS pixel is retrieved as specified in 3.5 with the DEM atlas.

#### 4.4 Linear interpolation

The linear interpolation of a function  $y(x)$  at point  $x$  from the function values  $y_0$  and  $y_1$  at points  $x_0$  and  $x_1$  writes:

$$y = y_0 + (y_1 - y_0) \frac{(x - x_0)}{(x_1 - x_0)} \quad \text{eq( 88)}$$

#### 4.5 Bilinear interpolation

The bi-linear interpolation of a function  $f(x,y)$  at the point  $(x;y)$  from the function values  $f_{00}$ ,  $f_{01}$ ,  $f_{11}$  and  $f_{10}$  at points  $(x_0; y_0)$ ,  $(x_0; y_1)$ ,  $(x_1; y_1)$ ,  $(x_1; y_0)$  on a regular 2D-grid writes:

$$f(x, y) = \alpha_{00} (x_1 - x)(y_1 - y) + \\ \alpha_{10} (x - x_0)(y_1 - y) + \\ \alpha_{01} (x_1 - x)(y - y_0) + \\ \alpha_{11} (x - x_0)(y - y_0) \quad \text{eq( 89)}$$

Where  $\alpha_{ij} = \frac{f(x_i, y_j)}{(x_1 - x_0)(y_1 - y_0)}$

#### 4.6 Euclidean norm

The Euclidean norm of a vector  $\vec{V} = [v_1, v_2, \dots, v_n]$  of dimensionn, noted  $\|\vec{V}\|$ , writes:

$$\|\vec{V}\| = \sqrt{\sum_{i=1}^n v_i^2} \quad \text{eq( 90)}$$

#### 4.7 Dot product

The dot product, also called scalar product, of two vectors  $\vec{U} = [u_1, u_2, \dots, u_n]$  and  $\vec{V} = [v_1, v_2, \dots, v_n]$  of dimension n writes:

$$\vec{U} \cdot \vec{V} = \sum_{i=1}^n u_i v_i \quad \text{eq( 91)}$$

#### 4.8 Vector product

The vector product, also called cross product, of two vectors  $\vec{V}_1$  and  $\vec{V}_2$  in Cartesian coordinates (x,y,z) writes:

$$\vec{V}_1 \times \vec{V}_2 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \times \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} y_1 z_2 - y_2 z_1 \\ z_1 x_2 - z_2 x_1 \\ x_1 y_2 - x_2 y_1 \end{pmatrix} \quad eq(92)$$

#### 4.9 Water-vapour density definitions and relationships

The humidity concentrations can be expressed in different units. We detail here the representations used in various functions of the processor and detail their relationships.

##### The water-vapour pressure, $p_{H_2O}$

The vapour pressure of moist air is defined as the partial pressure of the water-vapour present in the air mass. It is said to be with respect to liquid (or ice) water if the air mass is over a plane surface of liquid (or ice) water at the same temperature and pressure.

##### The water-vapour saturation pressure, $e_S$

The water-vapour saturation pressure is defined as the water-vapour pressure where two phases of water co-exist in neutral equilibrium. It is said to be with respect to liquid (or ice) water if the air mass is over a plane surface of liquid (or ice) water at the same temperature and pressure. The saturation water-vapour pressure,  $e_S$ , is given by the integrated Clausius Clapeyron equation, as formulated by the Goff Gratch equations [SmithMet].

Over water, the saturation water-vapour pressure in hPa is as follows:

$$\log_{10}(e_S) = -7.90298 * \left( \frac{T_{WB}}{T} - 1 \right) + 5.02808 * \log_{10} \left( \frac{T_{WB}}{T} \right) \\ - 1.3816 * 10^{-7} * \left( 10^{11.344 * (1 - \frac{T}{T_{WB}})} - 1 \right) + \\ 8.1328 * 10^{-3} * \left( 10^{-3.49149 * (\frac{T_{WB}}{T} - 1)} - 1 \right) + \log_{10}(e_{wB}) \quad eq(93)$$

where  $T$  is the temperature in °K,  $T_{WB} = 373.16$  °K, and  $e_{wB} = 1013.246$  hPa.

Over ice, the saturation water vapour pressure in hPa is as follows:

$$\log_{10}(e_S) = -9.09718 * \left( \frac{T_0}{T} - 1 \right) - 3.56654 * \log_{10} \left( \frac{T_0}{T} \right) + 0.876793 * \left( 1 - \frac{T}{T_0} \right) \\ + \log_{10}(e_{w0}) \quad eq(94)$$

where  $T_0 = 273.15$  K, and  $e_{w0} = 6.1071$  hPa.

##### Water-vapour mass mixing ratio, $r_{H_2O}$

The mass mixing ratio is the ratio of the mass of a constituent over the dry air of the parcel. It is expressed in kg/kg and writes for water-vapour:

$$r_{H_2O} = \frac{M_{H_2O}}{M_L} \cdot \frac{n_{H_2O}}{n_L} \quad eq(95)$$

Where  $M_{H_2O} = 18.01534$  g/mol is the molar mass of water and  $M_L = 28.964$  g/mol is the molar mass of dry air.  $n_{H_2O}$  and  $n_L$  are the number density of water-vapour and dry-air, respectively.

### Specific humidity, $q$

The specific humidity  $q$  is the ratio of water-vapour mass to the air parcel's total mass (i.e. moist air). It is often approximated the mass mixing ratio. It is expressed in kg/kg and writes.

$$q = \frac{r_{H_2O}}{(1 + r_{H_2O})} \quad \text{eq(96)}$$

Conversely

$$r_{H_2O} = \frac{q}{(1 - q)} \quad \text{eq(97)}$$

### Water-vapour partial pressure, $p_{H_2O}$

Let  $p$  and  $p_{H_2O}$  be the total atmospheric pressure and the partial pressure of water-vapour in the air parcel, respectively. The partial pressure of water-vapour relates to the mass mixing ratio:

$$p_{H_2O} = \frac{r_{H_2O} \cdot p}{\frac{M_{H_2O}}{M_L} + r_{H_2O}} \quad \text{eq(98)}$$

And to the specific humidity

$$p_{H_2O} = \frac{q \cdot p}{\frac{M_{H_2O}}{M_L} + q(1 - \frac{M_{H_2O}}{M_L})} \sim \frac{q \cdot p}{0.622 + 0.378q} \quad \text{eq(99)}$$

Conversely

$$q \sim \frac{0.622 p_{H_2O}}{p - 0.378 p_{H_2O}} \quad \text{eq(100)}$$

### Water-vapour volume mixing ratio, $w$

Another representation of water-vapour density is the volume mixing ratio,  $w$ . It is usually expressed in parts per million by volume or  $ppmv$ . It relates to the mass mixing ratio  $r_{H_2O}$  (in kg/kg) as follows

$$w = 10^6 \cdot \frac{M_L}{M_{H_2O}} \cdot r_{H_2O} \quad \text{eq(101)}$$

and to the specific humidity:

$$w = 10^6 \cdot \frac{M_L}{M_{H_2O}} \cdot \frac{q}{1 - q} \quad \text{eq(102)}$$

### Dew point temperature, $T_{dew}$

The dew point temperature,  $T_{dew}$ , is the temperature to which a given air parcel with water-vapour pressure  $p_{H_2O}$  must be cooled at constant pressure and constant water-vapour content in order for saturation to occur. It can be approximated with the following formula, in Kelvin:

$$T_{dew} = \frac{T_n}{\frac{m}{\log(p_{H_2O}/A)} - 1} + 273.15 \quad \text{eq(103)}$$

Where  $T_n$ ,  $m$  and  $A$  are configurable parameters dependent on the atmospheric temperature and if the saturation occurs over liquid or ice phase [Vaisala 2013].

	A	m	Tn	max error	Temperature range
water	6.116441	7.591386	240.7263	0.083%	-20...+50°C
	6.004918	7.337936	229.3975	0.017%	+50...+100°C
	5.856548	7.27731	225.1033	0.003%	+100...+150°C
	6.002859	7.290361	227.1704	0.007%	+150...+200°C
	9.980622	7.388931	263.1239	0.395%	+200...+350°C
	6.089613	7.33502	230.3921	0.368%	0...+200°C
ice	6.114742	9.778707	273.1466	0.052%	-70...0°C

*Table 7: Constants for formula eq( 103)*

### The Relative Humidity, %RH

The relative humidity of an air-water mixture is defined as the ratio of the partial pressure of water-vapour present at a given temperature and pressure,  $p_{H_2O}$ , to the partial pressure given by of the water present at saturation for the given temperature and pressure,  $e_S$ . Thus the relative humidity of air is a function of both water content and temperature. It is expressed in % and writes as follows:

$$\%RH = 100 \cdot \frac{p_{H_2O}}{e_S} \quad \text{eq( 104)}$$

### 4.10 Virtual temperature

The virtual temperature  $T_v$  of a moist air parcel of temperature  $T$  with specific humidity  $q$  is the temperature at which a theoretical dry air parcel would have the same total pressure and density. It writes:

$$T_v = T(1 + 0.608q) \quad \text{eq( 105)}$$

With  $T$  expressed in K and  $q$  in kg/kg.

### 4.11 Check for Super-saturation of Water Vapour

The check for super-saturation requires the comparison of the retrieved water-vapour amount with its saturation value. For each pressure level of a given profile, the saturation humidity  $q_S$  is computed with the equation eq(100) using the saturation water-vapour pressure  $e_S$  from eq( 93 ) and eq( 94 ) as water-vapour pressure  $p_{H_2O}$ . If the actual specific humidity  $q$  is higher than  $q_S$ , then it is reset to  $q_S$  and the correction for super-saturation is recorded in a processing flag.

### 4.12 Moist air specific heat and gas constant

Let  $r_{H_2O}$  be the water-vapour mixing ratio at a given altitude, expressed in kg/kg, then the corresponding moist air specific heat,  $c_{p,m}$ , and moist air gas constant,  $R_m$ , write [Bolton 1980][Iribarne and Godson 1973]:

$$R_m = R_L(1 + 0.608 r_{H_2O}) \quad \text{eq( 106)}$$

$$c_{p,m} = c_p(1 + 0.887 r_{H_2O}) \quad \text{eq( 107)}$$

where  $R_L = 287.06 \text{ J}. \text{kg}^{-1}. \text{K}^{-1}$  is the specific gas constant for dry air and  $c_p = 1005.71 \text{ J}. \text{kg}^{-1}. \text{K}^{-1}$  is the specific heat of dry air at constant pressure.

#### 4.13 Atmospheric adiabatic lapse rates

**Dry adiabatic lapse rate:** In the absence of water-vapour condensation, the temperature  $T$  of an unsaturated air parcel changing pressure during an adiabatic process follows the Poisson's equation. Hence, the relationship between temperature  $T$  and pressure  $p$  at two distinct atmospheric levels  $i$  and  $j$  along the dry adiabats (illustrated in Figure 30) writes:

$$p_j = \frac{p_i}{\left(\frac{T_i}{T_j}\right)^{R_m/c_{p,m}}} \quad \text{eq( 108)}$$

and

$$T_i = T_j \left(\frac{p_i}{p_j}\right)^{R_m/c_{p,m}} \quad \text{eq( 109)}$$

With  $R_m$  and  $c_{p,m}$  the moist air gas constant and specific heat computed as described in §4.12.

The **potential temperature** is the temperature that an unsaturated parcel at temperature  $T$  and pressure  $p$  would attain if adiabatically brought at a standard pressure of 1000 hPa. It is written  $\theta$  and is computed after:

$$\theta = T \left(\frac{1000}{p}\right)^{R_m/c_{p,m}} \quad \text{eq( 110)}$$

The dry adiabats (e.g. on Figure 30) are the line of constant potential temperature.

**Moist adiabatic lapse rate:** when water-vapour condenses during an adiabatic ascent, latent heat is released within the air parcel as it rises. Therefore the temperature decreases less rapidly than the dry adiabatic lapse rate above. The saturated lapse rate significantly varies with the water-vapour content. Analogous to the potential temperature,  $\theta$ , which is preserved along the dry adiabat, the **equivalent potential temperature**,  $\theta_e$ , expresses a constant quantity along the moist adiabat. It can be approximated as [Bolton 1980]:

$$\theta_e = \theta \exp \left[ \left( \frac{3.376}{T_{LCL}} - 0.00254 \right) * 1000 * r_{H_2O} \left( 1 + 0.81 * 10^{-3} * r_{H_2O} \right) \right] \quad \text{eq( 111)}$$

With

- $T_{LCL}$  the temperature at the air parcel at lifting condensation level (see 4.15)
- $\theta$  and  $r_{H_2O}$  are respectively the potential temperature -eq( 110)- and water-vapour mixing ratio (in kg/kg) of the air parcel at the successive levels.

#### 4.14 Check for Super-Adiabatic Layering

The lapse rate  $T_i/T_{i+1}$  between adjacent pressure levels  $p_i$  and  $p_{i+1}$  in the retrieved profiles (with  $p_{i+1} < p_i$ ) is compared to the dry adiabatic lapse  $b_i = T_i^{\text{adia}}/T_{i+1}^{\text{adia}}$  evaluated following equation eq( 109) in §4.13.

If in the retrieved profile the ratio  $T_i/T_{i+1} < b_i$ , then a correction term has to be calculated according to the following:

$$a_i = \frac{b_i T_{i+1} - T_i}{1 + b_i} \quad \text{eq( 112)}$$

If  $a$  is greater than the standard deviation (error estimates) of the temperature, then  $T_i$  has to be increased by  $a_i$  and  $T_1$  must be decreased by  $a$ . The correction is reported in a processing flag.

#### 4.15 Lifting condensation level (LCL)

The vertical lapse rate for an unsaturated air particle theoretically lifted is the dry adiabatic lapse rate (solid line in Figure 30) until the lifting condensation level (LCL) is reached.

The **temperature at LCL**,  $T_{LCL}$ , is computed first the empirical formula [Bolton 1980] (Note: parameters are configurable):

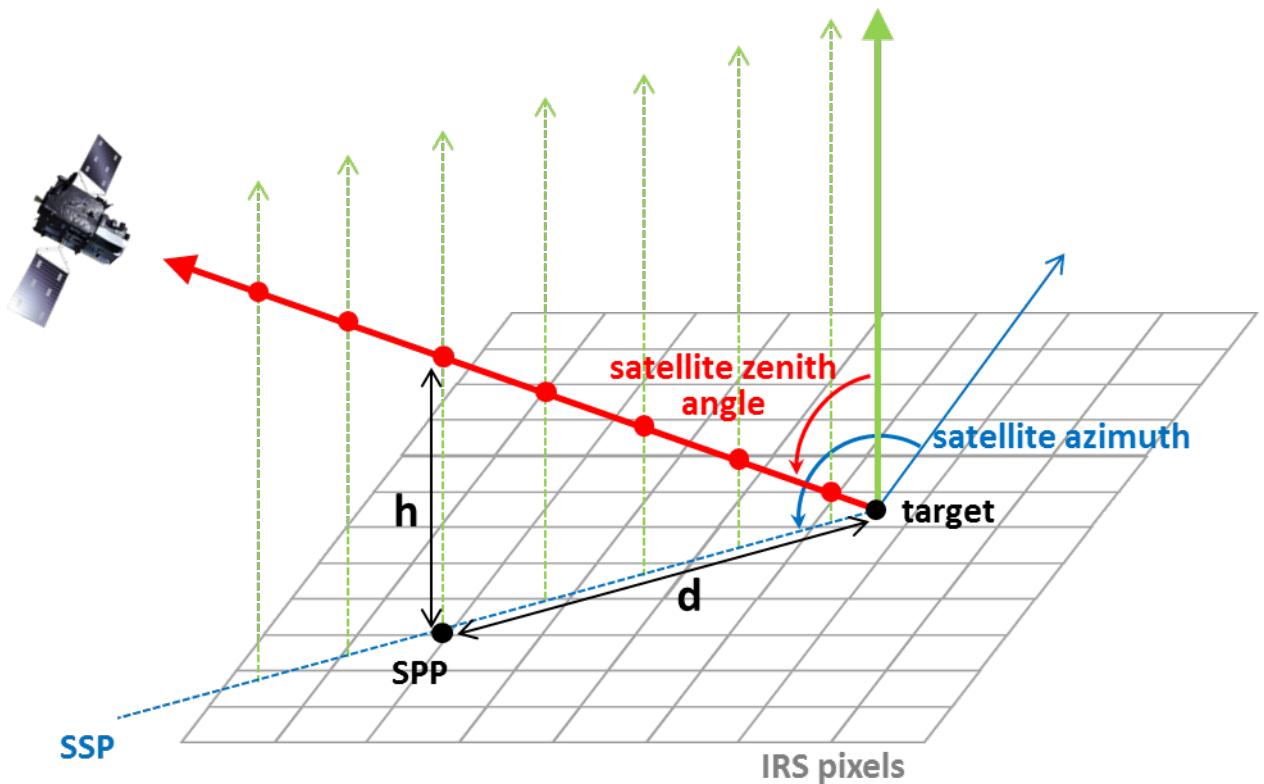
$$T_{LCL} = \frac{1}{\frac{1}{T_{parcel}} - \frac{\ln(RH_{parcel}/100)}{2840}} \quad \text{eq(113)}$$

where  $RH_{parcel}$  is the relative humidity of this air parcel at origin (usually at surface), expressed in %, derived from its origin temperature and pressure (usually at surface),  $T_{parcel}$  and  $p_{parcel}$ , as described in section 4.9.

The **pressure at LCL**,  $p_{LCL}$ , is the pressure of the lifted air parcel whose temperature decreased from  $T_0$  to  $T_{LCL}$  during the theoretical dry adiabatic ascent. It can thus be obtained from eq(108) of §4.13.

#### 4.16 Slanted levels geolocation

The purpose of this function is to compute the longitude/latitude at the vertical of discrete levels of a slanted profile (sub-profile points, SPP, see Figure 31), knowing their altitudes, the target longitude and latitude (assumed on a WGS84 ellipsoid) and given the target to satellite azimuth angle, the target satellite zenith angle and the reference altitude on target.



*Figure 31: Viewing geometry of target and sub-profile points*

Let

- $lon_{target}, lat_{target}$  be respectively the longitude and latitude of the IRS pixel centre (target)
- $\hat{z}_{target}, \widehat{Az}_{target}$  be the satellite zenith and azimuth angles at target, respectively
- $h$  be the reference altitude of the discrete level of the slanted profile
- $d$  the distance between the sub-profile point (SPP) and the target (pixel centre)
- $lon_{SPP}, lat_{SPP}$  be respectively the longitude and latitude of the sub-profile point
- $\widehat{Az}_{SPP}$  be the azimuth of the geodesic at the sub-profile point

Note: the sub-satellite point, SSP, does not play a role here and is not to be confused with the sub-profile point SPP.

Then we can write:

$$\theta = 90 - \hat{z}_{target} \quad eq(114)$$

$$d = h / \tan \theta \quad eq(115)$$

The calculation of the geo-coordinates of SPP is implemented via Vicenty's method [Vincenty], which describes a geodetic method to compute the location of a point that is at a given (ellipsoidal) distance and initial azimuth (direction) from another point.

A detailed derivation of Vicenty's method can, for instance, be found in [Vincenty]. In the below only the final algorithm is shown, with the limitation that the distance between the target and the sub-profile points along the surface of the ellipsoid,  $s$ , is approximated by  $d$ , the distance measured in a plane normal to the target-SPP vector in an Earth fixed frame.

WGS84 ellipsoid constants:

$$a = 6378137 \text{ m; major semi - axes of the WGS84 ellipsoid} \quad \text{eq( 116)}$$

$$b = 6356752 \text{ m; minor semi - axes of the WGS84 ellipsoid} \quad \text{eq( 117)}$$

$$f = (a - b)/a \quad \text{eq( 118)}$$

Where  $f$  is the Earth flattening.

The first part of the algorithm writes:

$$\tan U_1 = (1 - f) \cdot \tan(lat_{target}) \quad \text{eq( 119)}$$

$$\cos U_1 = 1/\sqrt{1 + \tan^2 U_1} \quad \text{eq( 120)}$$

$$\sin U_1 = \tan U_1 \cdot \cos U_1 \quad \text{eq( 121)}$$

$$\sigma_1 = \text{atan}(\tan U_1 \cdot \cos \widehat{Az}_{target}) \quad \text{eq( 122)}$$

$$\sin \alpha = \cos U_1 \cdot \sin \widehat{Az}_{target} \quad \text{eq( 123)}$$

$$\cos^2 \alpha = 1 - \sin^2 \alpha \quad \text{eq( 124)}$$

$$u^2 = \cos^2 \alpha \cdot (a^2 - b^2)/b^2 \quad \text{eq( 125)}$$

$$A = 1 + \frac{u^2}{16384} \cdot (4096 + u^2 \cdot (-768 + u^2 \cdot (320 - 175u^2))) \quad \text{eq( 126)}$$

$$B = \frac{u^2}{1024} \cdot (256 + u^2 \cdot (-128 + u^2 \cdot (74 - 47u^2))) \quad \text{eq( 127)}$$

$$\sigma = \frac{s}{bA}; s \sim d \quad \text{eq( 128)}$$

Then the following sequence is iterated until the change in  $\sigma$  is negligible (e.g.  $\Delta\sigma < 10^{-12}$ ).

{

$$\cos(2\sigma_m) = \cos(2\sigma_1 + \sigma) \quad \text{eq( 129)}$$

$$\begin{aligned} \Delta\sigma = & \left( \cos(2\sigma_m) \right. \\ & + \frac{B}{4} \left[ \cos \sigma \cdot (-1 + 2 \cos^2(2\sigma_m)) \right. \\ & \left. \left. - \frac{B}{6} \cdot \cos(2\sigma_m) \cdot (-3 + 4 \sin^2 \sigma) \cdot (-3 + 4 \cos^2(2\sigma_m)) \right] \right) \\ & \sigma += \Delta\sigma \end{aligned} \quad \text{eq( 130)}$$

}

$$lat_{SPP} = \text{atan} \left( \frac{\sin U_1 \cdot \cos \sigma + \cos U_1 \cdot \sin \sigma \cdot \cos \widehat{Az}_{target}}{(1 - f) \sqrt{\sin^2 \alpha + (\sin U_1 \cdot \sin \sigma \cdot \cos \widehat{Az}_{target})^2}} \right) \quad \text{eq( 132)}$$

$$\lambda = \text{atan} \left( \sin \sigma \cdot \frac{\sin \widehat{Az}_{target}}{\cos U_1} \cdot \cos \sigma - \sin U_1 \cdot \sin \sigma \cdot \cos \widehat{Az}_{target} \right) \quad \text{eq( 133)}$$

$$C = \frac{f}{16} \cdot \cos^2 \alpha \cdot (4 + f(4 - 3 \cos^2 \alpha)) \quad \text{eq( 134)}$$

$$\begin{aligned} L = & \lambda - (1 - C) \cdot f \cdot \sin \alpha \cdot \left( \sigma \right. \\ & \left. + C \sin \sigma (\cos(2\sigma_m) + C \cos \sigma (-1 + 2 \cos^2(2\sigma_m))) \right) \end{aligned} \quad \text{eq( 135)}$$

$$lon_{SPP} = lon_{target} + L \quad eq(136)$$

$$\widehat{Az}_{SPP} = \text{atan} \left( -\frac{\sin \alpha}{\sin U_1 \cdot \sin \sigma - \cos U_1 \cdot \cos \sigma \cdot \cos \widehat{Az}_{target}} \right) \quad eq(137)$$

An example of a C++ implementation of the above algorithm is given in Appendix E.

## APPENDIX A ASSUMPTIONS AND OPEN ISSUES

List of open issues, requiring prototyping or studies to define practical implementation (e.g. PWLR<sup>3</sup> window), characterise limitations in performances/scope of application of a given algorithm, define new or evolve existing algorithms etc.

ID Open issues	Short description	Scope
OI-1. PWLR <sup>3</sup> window size	The strategy for the PWLR <sup>3</sup> (e.g. spatial extent) has been defined consistently with IASI-NG specifications. Adjustments might be needed as a result of prototyping activities to account for border effects, computation time, configuration data size.	Day-2
OI-2. Instability indices	Based on initial requirements and MAG review, a set of instability indices shall be computed centrally and disseminated to the Users. Place-holders and samples are included in the product format description and in the algorithm functional blocks. A first set of algorithms was described to fulfil the needs of the L2PF price conversion –reflecting encoding and processing complexity. This selection has so far been confirmed from studies and interactions with users. The ATBD may evolve as the functions will be prototyped and evaluated for speed to optimise the processing. The list of instability indices and associated algorithms may further evolve as more experience is made of using hyperspectral sounding products for nowcasting.	Day-2
OI-3. Rim sounding	The algorithms described for IRS L2 Day-1 baseline have been demonstrated with IASI, which does not cover the full extent of IRS viewing geometry. Experience need to be made with sounding in the outer ring of the Earth disk, i.e. in the high to quasi-limb viewing mode. Specific new configuration or dedicated new algorithms may be needed to support applications in these viewing angles.	Day-2
OI-4. A priori slanted inputs	Systematic exhaustive evaluation of the geometry to reconstruct slanted first-guess/prior profiles from NWP forecasts may introduce too large computation overhead wrt required timeliness. Simplification or static solutions with look-up tables may need to be considered.	Day-1
OI-5. Slant to vertical, pre-/post-processing	End-products are required at vertical profiles while the observations are slanted. In first instance, the retrievals are described along the slant path and post-processed to re-construct vertical profiles. An approach with configurable look-up table is proposed as an alternative to systematic online collocations. This will need to be revisited with instrument/scanning design if the assumption holds, e.g. how much adjacent dwell relative mapping can vary.	Day-1

OI-6. Dwell staging	Meanwhile, studies will be performed to possibly directly retrieve the vertical profiles from slanted observations, which may require staging the inputs L1 instead of the L2 outputs.	Day-2
OI-7. Potential for air quality and atmospheric composition products	Besides O <sub>3</sub> retrievals and a pseudo-quantitative dust indicator, no atmospheric composition products are regarded at Day-1. Potential for detection and quantitative retrieval of e.g. CO, NH <sub>3</sub> , SO <sub>2</sub> and ash/dust composition should be evaluated as soon as the Day-1 baseline is consolidated, aiming operational service shortly after end of commissioning.	Day-2
OI-8. Dynamic background error for forecast <i>a priori</i>	In the option where forecast profiles are used as <i>a priori</i> in the OEM, studies are needed to determine the optimal configuration of the background error. This can be achieved by using for instance the error covariance estimated from the ensemble model prediction at ECMWF. The study will address possible pre-processing to this error estimate as well as practical aspects of the data stream for near-real time processing.	Day-2
OI-9. Synergetic cloud mask with FCI	A Day-2 possible development for IRS L1 processing includes use of FCI cloud mask, to complement the IRS radiances with some collocation cloud information. This external information could be valuable for the L2 processing itself, either for simple cloud screening or as additional input information.	Day-2

**APPENDIX B**
**REFERENCES**

ID	Reference
[Aires et al. 2002]	Aires et al., “A regularised neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument”, <i>J. Appl. Meteorology</i> , 41:144-159, February 2002
[Atkins and Wakimoto 1991]	Atkins, N.T., és R.M. Wakimoto, 1991: Wet microburst activity over the Southeastern United States: Implications for forecasting. <i>Wea. Forecasting</i> , 6, 470-482.
[ACE-2]	Berry et al., “ACE2: The New Global Digital Elevation Model”, in “Gravity, Geoid and Earth Observation”, pp 231-237, Springer 2010, DOI: 10.1007/978-3-642-10634-7_30
[Armante et al. 2013]	EUMETSAT Study 2013/205792 – “Characterisation of an observation error for IASI using collocated sondes”, Armante et al. 2013, Laboratoire de Météorologie Dynamique (LMD). Final reports DOCS#706293, DOCS#968203.
[August et al. 2012]	August et al. “IASI on Metop-A: Operational Level 2 retrievals after five years in orbit”, <i>Journal of Quantitative Spectroscopy and Radiative Transfer</i> , 2012, Volume 113, Issue 11, p. 1340-1371 – DOI: 10.1016/j.jqsrt.2012.02.028
[Aumann et al. 2003]	Aumann et al., “AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products, and Processing Systems”, <i>IEEE Transactions On Geoscience And Remote Sensing</i> , Vol. 41, No. 2, February 2003
[Blackwell et al. 2005]	Blackwell W.J., “A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectralresolution sounding data”, 2005, <i>IEEE Trans. Geosci. Remote Sens.</i> , 43(11), 2535–2546
[Blumstein et al. 2004]	Blumstein et al., 2004, “IASI instrument: technical overview and measured performances”. <i>Proceedings of the SPIE. Infrared Spaceborne Remote Sensing XII</i> . 5543: 196–207
[Bolton 1980]	Bolton, D., 1980, “The computation of equivalent potential temperature”, <i>Monthly Weather Review</i> , 108 (7), 1046{1053.
[Bormann 2016]	Bormann N., “Slant path radiative transfer for the assimilation of sounder radiances”, <i>ECMWF Technical Memo 782</i>
[Boylan et al. 2016]	Boylan et al., “Identification and intercomparison of surface-based inversions over Antarctica from IASI, ERA-Interim, and Concordiasi dropsonde data”, <i>JGR</i> 2016
[Bryan 2008]	Bryan, 2008, “On the Computation of Pseudoadiabatic Entropy and Equivalent Potential Temperature”, <i>AMS Notes and Correspondence</i> , DOI: 10.1175/2008MWR2593.1
[CAMEL]	Borbas et al., “NASA MEASUREs Combined ASTER and MODIS Emissivity over Land (CAMEL)”, American Geophysical Union, Fall General Assembly 2016, abstract #GC51D-1203  <a href="http://cimss.ssec.wisc.edu/iremis/">http://cimss.ssec.wisc.edu/iremis/</a> <u>Papers in preparation</u>
[CAMS]	Copernicus Atmosphere Monitoring Service <a href="https://www.ecmwf.int/en/about/what-we-do/84opernicus/84opernicus-atmosphere-monitoring-service">https://www.ecmwf.int/en/about/what-we-do/84opernicus/84opernicus-atmosphere-monitoring-service</a>
[Capelle et al. 2011]	Capelle V. et al., “Infrared Continental Surface Emissivity Spectra and Skin Temperature Retrieved from IASI Observations over the Tropics”, 2011, <i>Journal Of Applied Meteorology And Climatology</i> , doi: 10.1175/JAMC-D-11-0145.1
[Chevallier 2002]	Chevallier, “Sampled databases of 60-level atmosoheric profiles fom the ECMWF analyses”, NWPSAF-EC-TR-004
[Chylek et al. 2006]	Chylek et al., “Comparison of near-infrared and thermal infrared cloud phase detections”, 2006, <i>JGR</i> , doi:10.1029/2006JD007140
[Clarisse et al. 2012]	Clarisse et al., 2012, “Retrieval of sulphur dioxide from the infrared atmospheric sounding interferometer (IASI)”, <i>Atmos. Meas. Tech.</i> , doi: 10.5194/amt-5-581-2012
[Clarisse et al. 2013]	Clarisse et al., 2013, “A unified approach to infrared aerosol remote sensing and type specification”, <i>Atmospheric Chemistry and Physics</i> , 2013., doi:10.5194/acp-13-2195-2013
[Collard and Matricardi 2005]	Collard and Matricardi, “Definition of an efficient interface to NWP for assimilation radiances”, EUM WP 989-2 December 2005

<b>ID</b>	<b>Reference</b>
[Crapeau <i>et al.</i> 2017]	Crapeau et al. 2017, "Experimenting different a priori sources for optimal estimation retrievals with IASI", ITSC-21, December 2017, Darmstadt, Germany.
[Crevoisier <i>et al.</i> 2009]	Crevoisier et al., 'First year of upper tropospheric integrated content of CO <sub>2</sub> from IASI hyperspectral infrared observations', Atmos. Chem. Phys., 2009, 9, 4797–4810, doi:10.5194/acp-9-4797-2009
[De Souza <i>et al.</i> 2006]	De Souza et al., "Infrared dust spectral signatures from AIRS", 2006, Geophysical Research Letters, Vol. 33, L03801, Doi:10.1029/2005GL024364
[Dee 2004]	Dee, D. P., (2004) "Variational bias correction of radiance data in the ECMWF system.", Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP. Reading, UK, 28 June – 1 July 2004. Available from: <a href="http://www.ecmwf.int/publications">http://www.ecmwf.int/publications</a>
[Doswell <i>et al</i> 1982]	Doswell, C. A., III, J. T. Schaefer, D. W. McCann, T. W. Schlatter, and H. B. Wobus, 1982: "Thermodynamic analysis procedures, at the National Severe Storms Forecast Center". Preprints, Ninth Conf. on Weather Forecasting and Analysis, Seattle, WA, Amer. Meteor. Soc., 304–309.
[EARS IASI L2]	EUMETSAT Advanced Retransmission System IASI L2 <a href="https://www.eumetsat.int/website/home/News/DAT_3709776.html">https://www.eumetsat.int/website/home/News/DAT_3709776.html</a> <a href="https://www.eumetsat.int/website/home/Data/RegionalDataServiceEARS/EARSIASI/index.html">https://www.eumetsat.int/website/home/Data/RegionalDataServiceEARS/EARSIASI/index.html</a>
[ECMWFgrid137]	<a href="https://www.ecmwf.int/en/forecasts/documentation-and-support/137-model-levels">https://www.ecmwf.int/en/forecasts/documentation-and-support/137-model-levels</a>
[ERA-5]	<a href="http://climate.copernicus.eu/products/climate-reanalysis">http://climate.copernicus.eu/products/climate-reanalysis</a>
[Eyre 1992]	Eyre, J. R., "A bias correction scheme for simulated TOVS brightness temperatures. ECMWF, 1992, Technical Memorandum 186. Available from: <a href="http://www.ecmwf.int/publications/">http://www.ecmwf.int/publications/</a>
[Eyre and Menzel 1989]	Eyre, J. R., and W. P. Menzel, 1989, "Retrieval of cloud parameters from satellite sounder data: A simulation study." J. Appl. Meteor., 28, 267–275. Doi: 10.1175/1520-0450(1989)028<0267:ROCPFS>2.0.CO;2
[Eyre <i>et al.</i> 2011]	J. Eyre, N. Bormann, E. Pavelin, F. Rabier, L. Stewart, " Elaboration of user requirements for MTG-IRS products and implications for data processing methods: a discussion paper", 1 <sup>st</sup> April 2011
[FCI L2 PS]	MTG-FCI Level 2 Processing Specification, EUM/MTG/SPE/12/1078 v3F
[Feltz <i>et al.</i> 2017]	Feltz et al., "Assessment of NOAA NuCAPS upper air temperature profiles using COSMIC GPS radio occultation and ARM radiosondes", JGR 2017, DOI: 10.1002/2017JD026504
[Galway 1956]	J. G. Galway, 1956: "The Lifted Index as a Predictor of Latent Instability," Bulletin of the American Meteorological Society, Vol. 37, 528-529
[George 1960]	J.J. George (1960): Weather Forecasting for Aeronautics. New York City: Academic Press. P. 673
[Goukenleuque <i>et al.</i> 2017]	Goukenleuque et al., "Preparing test data for the IRS Level 2 processor", 21 <sup>st</sup> International TOVS Study Conference, Darmstadt, Germany, 2017
[Gray and Pavelin 2017]	Gray et Pavelin, "Development of a Dynamic Infrared Land Surface Emissivity Atlas based on IASI Retrievals", 21 <sup>st</sup> International TOVS Study Conference, Darmstadt, Germany, 2017
[Hadji-Lazaro <i>et al.</i> 1999]	Lazaro et Clerbaux, "An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances", Journal Of Geophysical Research, Vol. 104, No. D19, Pages 23,841-23,854, October 20, 1999
[Haklander and Delden 2003]	Haklander, A.J., Van Delden, A., 2003, "Thunderstorm predictors and their forecast skill for the Netherlands", Atmospheric Research, Vol. 67-68, July/September 2003, pp 273-299
[Han <i>et al.</i> 2013]	Han et al., , 2013 "Suomi NPP CrIS measurements, sensor data record algorithm,calibration and validation activities, and record data quality", Journal Of Geophysical Research: Atmospheres, Vol. 118, 12,734–12,748, doi:10.1002/2013JD020344

<b>ID</b>	<b>Reference</b>
[Han and Sohn 2013]	Han and Sohn, "Retrieving Asian dust AOT and height from hyperspectralsounder measurements: An artificial neural network approach", Journal of Geophysical Research: Atmospheres, VOL. 118, 837–845, Doi:10.1002/Jgrd.50170, 2013
[Huang et al. 2004]	Huang et al. "Minimum Local Variance Retrieval of Cloud Altitude and Effective Spectral Emissivity – Simulation and Initial Verification" J. of Applied Meteorology, Volume 43, May 2004
[Hultberg and August 2016]	Hultberg T. and August T., "What is the error?", 4 <sup>th</sup> IASI Conference 2016
[Hultberg and August 2014]	Hultberg and August, "Use of reconstructed radiances in EUMETSAT's IASI Level 2 processor", ITSC- XIX 2014
[Hultberg and August 2015]	Hultberg and August, "Evolutions and self-organisation of the piecewise linear regression for IASI", ITSC-XX 2015
[König et al. 2001]	König, M., Tjemkes, S. A., Kerkmann, J., 2001: Atmospheric Instability Parameters Derived from MSG SEVIRI Observations. 11 <sup>th</sup> Conference on Satellite Meteorology and Oceanography
[Kocsis et al., 2017]	Final report: "Investigation of MSG SEVIRI and EPS IASI derived atmospheric instability in relation of other observations" EUMETSAT Study EUM/CO16/4600001802/KJG
[Kunz 2007]	Kunz M., 2007, "The skill of convective parameters and indices to predict isolated and severe thunderstorms", Nat. Hazards Earth Syst. Sci., 7, 327–342, 2007
[IASI L2 PGS]	IASI Level 2 Product Generation Specifications, EPS.SYS.SPE.990013
[IASI L2 v6 val]	IASI L2 v6 validation report, EUM/TSS/REP/14/776443
[IASI L2 v6.2 val]	IASI L2 v6.2 validation report, EUM/RSP/REP/16/857500
[IASI L2 v6.4 val]	IASI L2 v6.4 validation report, EUM/RSP/REP/16/857500
[IASI PCC val]	"EPS Product Validation Report: IASI L1 PCC PPF", Part. 1: EUM/OPS-EPS/REP/10/0148 , Part 2: EUM/OPS-EPS/REP/11/0036
[IASI-B L2val]	Metop-B/IASI L2 validation report, EUM/TSS/REP/13/684650, 172pp
[Iribarne and Godson 1973]	Iribarne J. V. and Godson W. L., 1973, "Atmospheric Thermodynamics", Geophysics and Astrophysics monographs, vol. 6, Reidel, 222pp.
[IRS MAG 5]	Minutes of the 5 <sup>th</sup> MTG-IRS Mission Advisory Group meeting, EUM/RSP/MIN/18/997984
[Jang et al. 2017]	"Improved AIRS Temperature and Moisture Soundings with Local A Priori Information for the 1DVAR Method", Journal Of Atmospheric And Oceanic Technology 2017, DOI: 10.1175/JTECH-D-16-0186.1
[Le Cun et al. 1998]	Le Cun et al., "Efficient backprop" in "Neural Networks: tricks of the trade", Springer, 1998
[Lorenz 1955]	Lorenz, 1955, "Available potential energy and the maintenance of the general circulation", Tellus 7(2): 157–167, doi:10.1111/j.2153-3490.1955.tb01148.x.
[MacQueen]	J.B.MacQueen, "Some methods for classification and Analysis of multivariate observations", Proceedings of 5 <sup>th</sup> Berkeley Symposium on Mathematical Statistics and Probability, 1967
[Matricardi 2014]	Matricardi et McNally, 2014, "The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var and the generation of RTTOV and PC_RTTOV new regression coefficients for AIRS, IASI, and IASI-NG", final report EUMETSAT study EUM/CO/07/4600001011/PS
[Matricardi 2017]	Matricardi et McNally, 2017, "The extension of PC_RTTOV to variable trace gas and aerosol simulations and the use of reconstructed radiances to assimilate the full IASI spectrum", final report EUMETSAT study EUM/CO/14/4600001332/TA
[McNally and Watts 2003]	McNally T. and Watts Ph., "A cloud detection algorithm for high-spectral-resolution infrared sounders", 2003, Quarterly Journal of the Royal Meteorological Society, doi: 10.1256/qj.02.208
[Milstein Blackwell 2016]	Miltsein et Blackwell, "Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS", JGR 2016, DOI: 10.1002/2015JD024008
[MTG PP]	MTG Programme Proposal – EUM/C/69/10/DOC/02
[NuCAPS ATBD]	The NOAA Unique CrIS/ATMS Processing System (NuCAPS): Algorithm Theoretical Basis Documentation

<b>ID</b>	<b>Reference</b>
[Pavelin and Candy 2013]	Pavelin EG, Candy B. "Assimilation of surface-sensitive infra-red radiances over land: estimation of land surface temperature and emissivity", Q J R Meteorol Soc, 140, 1198-1208 (2013), DOI:10.1002/qj.2218
[Putsay <i>et al.</i> 2017]	Putsay M, Kocsis S., Simon A., "Investigation of MSG SEVIRI and EPS IASI derived atmospheric instability in relation of other observations", Final Report on EUMETSAT study EUM/CO16/4600001802/KJG.
[Reichler et al. 2003]	Reichler et al. 2003, "Determining the tropopause height from gridded data", GRL 2003, doi:10.1029/2003GL018240
[Rodgers 2000]	Rodgers C., "Inverse methods for atmospheric sounding", World Scientific 2000
[Roman <i>et al.</i> 2016]	Roman et al., "A global assessment of NASA AIRS v6 and EUMETSAT IASI v6 precipitable water vapor using ground-based GPS SuomiNet stations", JGR 2016, DOI: 10.1002/2016JD024806
[RTTOV 12]	RTTOV v12 Users Guide, NWPSAF-MO-UD-037 <a href="https://nwpsaf.eu/site/download/documentation/rtm/docs_rttov12/users_guide_rttov12_v1.1.pdf">https://nwpsaf.eu/site/download/documentation/rtm/docs_rttov12/users_guide_rttov12_v1.1.pdf</a>
[Smith and Frey 1990]	W.L. Smith and R. Frey: "On cloud altitude determinations from High Resolution Interferometer Sounder (HIS) Observations" J. of Applied Meteorology, Volume 29, 1990
[Smith <i>et al.</i> 2012]	Smith et al., "Dual-Regression Retrieval Algorithm for Real-Time Processing of Satellite Ultraspectral Radiances", JAMC 2012, doi: 10.1175/JAMC-D-11-0173.1
[Smith and Woolf 1976 ]	Smith, W.L. and H.M. Woolf, "The use of eigenvectors of statistical co-variance matrices for interpreting satellite sounding radiometer observations.", J. Atmos. Sci., 33, 1127–1140
[SmithMet]	R.J. List, 1958: Smithsonian Meteorological Table, 6 <sup>th</sup> Edition, Smithsonian, Institution Press, Washington D.C.
[Sokolik 2002]	Sokolik I., "The spectral radiative signature of wind-blown mineral dust: Implications for remote sensing in the thermal IR region", 2002, Geophysical Research Letters, Vol. 29, No. 24, 2154, doi:10.1029/2002GL015910
[Strabala and Menzel 1994]	Strabala K. and Menzel P., "Cloud properties inferred from 8-12 $\mu$ m data", 1994, Journal of Applied Meteorology
[Strang]	Gilbert Strang, "Introduction to Linear Algebra", Wellesley-Cambridge Press
[Stubenrauch <i>et al.</i> 1999]	C. Stubenrauch et al, "Clouds as Seen by satellite, Sounders (3I) and Imagers (UISCCP). Part II: A, New Approach for Cloud Parameter, Determination in the 3I Algorithms", Journal of Climate 1999, <a href="https://doi.org/10.1175/1520-0442(1999)012&lt;2214:CASEBSS&gt;2.0.CO;2">https://doi.org/10.1175/1520-0442(1999)012&lt;2214:CASEBSS&gt;2.0.CO;2</a>
[Sun <i>et al.</i> 2017]	Sun B. et al, "Value of Dedicated, GRUAN and Conventional Radiosondes in Satellite Atmospheric Sounding Assessment", EUMETSAT User Conference 2017
[Susskind s. 2003]	Susskind et al., "Retrieval of Atmospheric and Surface Parameters From AIRS/AMSU/HSB Data in the Presence of Clouds", IEEE Transactions On Geoscience And Remote Sensing, Vol. 41, No. 2, February 2003
[Susskind <i>et al.</i> 2014]	Susskind et al. 2014, "Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: The atmospheric infrared sounder science team version-6 retrieval algorithm", DOI: 10.1111/1.JRS.8.084994
[Trent <i>et al.</i> 2016]	Trent T. et al., "Assessing The Impact of Aerosol on the Accuracy of IASI SST", 4 <sup>th</sup> IASI Conference 2016
[Vaisala 2013]	"Humidity conversion formulas", Vaisala 2013, B210973EN-F
[Vincenty]	Vincenty, T., 1975b, "Geodetic inverse solution between antipodal points", Technical report, DMAAC Geodetic Survey Squadron. doi:10.5281/zenodo.32999 Vincenty's method: <a href="https://en.wikipedia.org/wiki/Vincenty%27s_formulae">https://en.wikipedia.org/wiki/Vincenty%27s_formulae</a>
[Wang <i>et al.</i> 2016]	Wang et al., "Benefits of spatial resolution for next generation CrIS", 4 <sup>th</sup> IASI Conference, 2016. <a href="https://iasi.cnes.fr/sites/default/files/drupal/201612/default/bpc_iasi-conference4-s11-143_likun_wang.pdf">https://iasi.cnes.fr/sites/default/files/drupal/201612/default/bpc_iasi-conference4-s11-143_likun_wang.pdf</a>
[Whitburn <i>et al.</i> 2017]	Whitburn et al., "IASI-derived NH <sub>3</sub> enhancement ratios relative to CO for the tropical biomass burning regions", Atmos. Chem. Phys. Discuss. , 2017, doi:10.5194/acp-2017-331
[WMO 1957]	World Meteorological Organization (WMO), Meteorology A Three-Dimensional Science: Second Session of the Commission for Aerology, WMO

<b>ID</b>	<b>Reference</b>
	Bulletin IV(4), WMO, Geneva, 134– 138, 1957.
[Zhou et al. 2005]	Zhou et al., “Thermodynamic and cloud parameter retrieval using infrared spectral Data”, Geophysical Research Letters 2005, doi:10.1029/2005GL023211
[Zhou et al. 2007]	Zhou et al., “Physically Retrieving Cloud and Thermodynamic Parameters from Ultraspectral IR Measurements”, Journal Of The Atmospheric Sciences 2007, doi:10.1175/JAS3877.1
[Zhou et al. 2011]	Zhou et al., “Global Land Surface Emissivity Retrieved From Satellite Ultraspectral IR Measurements”, 2011, IEEE Transactions on Geoscience and Remote Sensing 49(4):1277 – 1290, DOI: 10.1109/TGRS.2010.2051036

**APPENDIX C            ACRONYMS**

<b>Acronym</b>	<b>Meaning</b>
ANN	Artificial Neural Network
API	Application Programming Interface
ATBD	Algorithm Theoretical Basis Document
ATOVS	Advanced TIROS Operational Vertical Sounder, EPS
AVHRR	Advanced Very High Resolution Radiometer, EPS
CAMEL	Combined ASTER and MODIS Emissivity over Land
CAMS	Copernicus Atmosphere Monitoring Service
CAPE	Convective Available Potential Energy
CFI	Customer Furnished Item
CIN	Convective INhibition
DEM	Digital Elevation Model
DMT	Document Management Tool
ECMWF	European Centre for Medium range Weather Forecasting
EOF	Empirical Orthogonal Function
EPS	EUMETSAT Polar System
EPS-SG	EUMETSAT Polar System – Second Generation
FCI	
FG	First Guess
FRTM	Fast Radiative Transfer Model
FTS	Fourier Transform Spectrometer
GHG	Green-house gases
GS	Ground Segment
H/W	Hardware
IASI	Infrared Atmospheric Sounder Interferometer, EPS
IASI-NG	Infrared Atmospheric Sounder Interferometer - Next Generation
IDEF-0	Integrated DEFinition 0 (function modelling method)
IR	Infrared
ISRF	Instrument Spectral Response Function
L1C	Level 1C
L2	Level 2
LCL	Lifting condensation level
LFC	Level of Free Convection
LI	Lifted-Index
LWIR	Long-Wave Infra-Red
MHS	Microwave Humidity Sounder (part of ATOVS), EPS
MLCAPE	Mixed-Layer CAPE
MLST	Mean Local Solar Time
MUCAPE	Maximum Unstable CAPE
MW	Micro-wave
MWIR	Mid-Wave Infra-Red
NIR	Near Infrared
NRT	Near Real Time
NWP	Numerical Weather Prediction
OSI-SAF	Ocean and Sea-Ice SAF
PCC	Principal Component Compression
PCR	Principal Component Residuals
PCS	Principal Component Scores
PS	Processing Specifications
Ps	Surface pressure
FS	Format Specifications
PPF	Product Processing Facility (EPS terminology for processing SW)

Acronym	Meaning
PSF	Point Spread Function
RO	Radio Occultation
SAD	Static Auxiliary Data
SAF	Satellite Application Facility
S/W	Software
TIR	Thermal Infrared
TRG	Trace Gases
Ts	Surface temperature
VIS	Visible (spectral range)

## APPENDIX D DRAFT PRODUCTS CONTENT TABLES

The IRS Level 2 products are organised in two product files:

- IRS-2-SVP: State Vector Product, contains the retrieved geophysical quantities and associated processing information, quality indicators
- IRS-2-COV: Contains the theoretical posterior Error Covariance Matrix, after the optimal estimation method.

The following tables give an indicative overview of the products content and sizes.

### D.1 IRS-2-SVP

DESCRIPTION	UNITS	DIM1	DIM2	DIM3	DIM4	DIM5	TYPE	SIZE
<b>General information</b>								
Contextual parameters: e.g. lat/lon, angles etc.		160	160	10	1	1	4	1024000 zlib compressed
Processing flags		160	160	10	1	1	2	512000 (ratio = 4)
<i>Sub-Total General information (MB)</i>							1.54	0.38
<b>Temperature and Water-Vapour</b>								
Processing flags		160	160	10	1	1	2	512000
<i>First retrieval (statistical)</i>								
Temperature profile	K	160	160	101	1	1	2	5171200
Water-vapour profile	kg/kg	160	160	101	1	1	4	10342400
Total column integrated water-vapour	kg.m <sup>-2</sup>	160	160	1	1	1	4	102400
Uncertainty estimates for temperature profile	K	160	160	40	1	1	2	2048000
Uncertainty estimates for water vapour profile	kg/kg	160	160	30	1	1	4	3072000
Near surface air temperature	K	160	160	1	1	1	2	51200
Near surface air water vapour	kg/kg	160	160	1	1	1	4	102400
<i>Second retrieval (OEM)</i>		160						
Temperature	K	160	160	101	1	1	2	5171200
Water vapour	kg/kg	160	160	101	1	1	4	10342400

DESCRIPTION	UNITS	DIM1	DIM2	DIM3	DIM4	DIM5	TYPE	SIZE
Total column integrated water-vapour	kg.m <sup>-2</sup>	160	160	1	1	1	4	102400
<i>Optimal Estimation Diagnostics</i>								
State vector cost of optimal estimation retrieval		160	160	1	1	1	4	102400
Measurement cost of optimal estimation retrieval		160	160	1	1	1	4	102400
Cost of first guess retrieval		160	160	1	1	1	4	102400
Number of optimal estimation iterations		160	160	1	1	1	4	102400
								zlib compressed
Instability indices (TBD)		160	160	10	1	1	4	1024000 (ratio = 4)
<i>Sub-Total TWV profiles (MB)</i>							38.45	9.61
<b>Ozone</b>								
Processing flags		160	160	10	1	1	2	512000
<i>First retrieval (Statistical)</i>								
First-guess ozone profile	kg/kg	160	160	101	1	1	4	10342400
Total column integrated ozone	kg.m <sup>-2</sup>	160	160	1	1	1	4	102400
Uncertainty estimates for ozone profile	kg/kg	160	160	20	1	1	4	2048000
<i>Second retrieval (OEM)</i>		160						
Ozone profile	K	160	160	101	1	1	4	10342400
Total column integrated ozone	kg.m <sup>-2</sup>	160	160	1	1	1	4	102400
<i>Optimal Estimation Diagnostics</i>								
State vector cost of optimal estimation retrieval		160	160	1	1	1	4	102400
Measurement cost of optimal estimation retrieval		160	160	1	1	1	4	102400
Cost of first guess retrieval		160	160	1	1	1	4	102400 (zlib compressed)
Number of optimal estimation iterations		160	160	1	1	1	4	102400 (ratio = 4)
<i>Sub-Total Ozone profiles (MB)</i>							23.86	5.96
<b>Surface parameters</b>								

DESCRIPTION	UNITS	DIM1	DIM2	DIM3	DIM4	DIM5	TYPE	SIZE
Processing flags		160	160	10	1	1	2	512000
<i>First retrieval (statistical) or a priori</i>								
Surface skin temperature (land and sea)	K	160	160	1	1	1	2	51200
Quality indicator for surface skin temperature	K	160	160	1	1	1	2	51200
Surface pressure	hPa	160	160	1	1	1	2	51200
Quality indicator for surface pressure	hPa	160	160	1	1	1	2	51200
Land surface emissivity in selected channels	-	160	160	10	1	1	2	512000
Quality indicator for land surface emissivity	-	160	160	1	1	1	2	51200
<i>Second retrieval (OEM)</i>								<i>zlib compressed</i>
Surface skin temperature	K	160	160	1	1	1	2	51200 <i>(ratio = 4)</i>
<i>Sub-Total Surface parameters (MB)</i>							<b>1.33</b>	<b>0.33</b>
<b>Cloud parameters</b>								
Processing flags		160	160	10	1	1	2	512000
Fractional cloud cover	%	160	160	2	1	1	1	51200
Cloud top pressure	hPa	160	160	2	1	1	2	102400
Cloud top temperature	K	160	160	2	1	1	2	102400
Predicted window channel Obs minus Calc	K	160	160	1	1	1	2	51200
Indicator of dust		160	160	1	1	1	1	25600 <i>zlib compressed</i>
Cloud phase		160	160	1	1	1	4	102400 <i>(ratio = 4)</i>
<i>Sub-Total Cloud parameters (MB)</i>							<b>0.95</b>	<b>0.24</b>
<b>Grand total size (MB / dwell)</b>							<b>66.12</b>	<b>16.53</b>

## D.2 IRS-2-COV

DESCRIPTION	UNITS	DIM1	DIM2	DIM3	DIM4	DIM5	TYPE	SIZE
<b>General information</b>								
Contextual parameters: e.g. lat/lon, angles etc.		160	160	10	1	1	4	1024000
Processing flags		160	160	10	1	1	2	512000
<b>Temperature and Water-Vapour</b>								
<i>Optimal Estimation Theoretical Error</i>								
Index of the error data record corresponding to the IFOVs in the line (=0xFFFF if N/A)		160	160	1	1	1	2	51200
Retrieval error covariance matrix for water vapour, compressed in principal component domain ( <i>in clear-sky only, 25% on average</i> )		6400	406	1	1	1	4	4377600
Retrieval error covariance matrix for temperature, compressed in principal component domain ( <i>in clear-sky only, 25% on average</i> )		6400	171	1	1	1	4	10393600
<b>Ozone</b>								
<i>Optimal Estimation Theoretical Error</i>								
Index of the error data record corresponding to the IFOVs in the line (=0xFFFF if N/A)		160	160	1	1	1	2	51200
Retrieval error covariance matrix for ozone, compressed in principal component domain ( <i>in clear-sky only, 25% on average</i> )		6400	55	1	1	1	4	1408000
								zlib compressed (ratio = 1)
<b>Grand Total TWV (MB / dwell)</b>	<b>17.82</b>	<b>17.82</b>						



## APPENDIX E RADIANCE BIAS CORRECTION

All retrieval methods involving forward model computations rely on the fact that the real observations made by the instrument (OBS) can be reproduced by synthetic calculations (CALC). This includes the statistical retrieval methods trained up-front with synthetic data (i.e. with an atmospheric climatology and their corresponding calculated radiances) and the variational retrieval (or assimilation) methods where the estimated atmospheric state vector is adjusted so that calculated spectra fit the observations .

The optimal estimation method specified in the IRS L2 ATBD is an example of the latter, where the cost function  $J$ , known to the readers, is minimised iteratively:

$$J = (x - x_a)^T \cdot S_x^{-1} \cdot (x - x_a) + (y - F(x))^T \cdot S_y^{-1} \cdot (y - F(x))$$

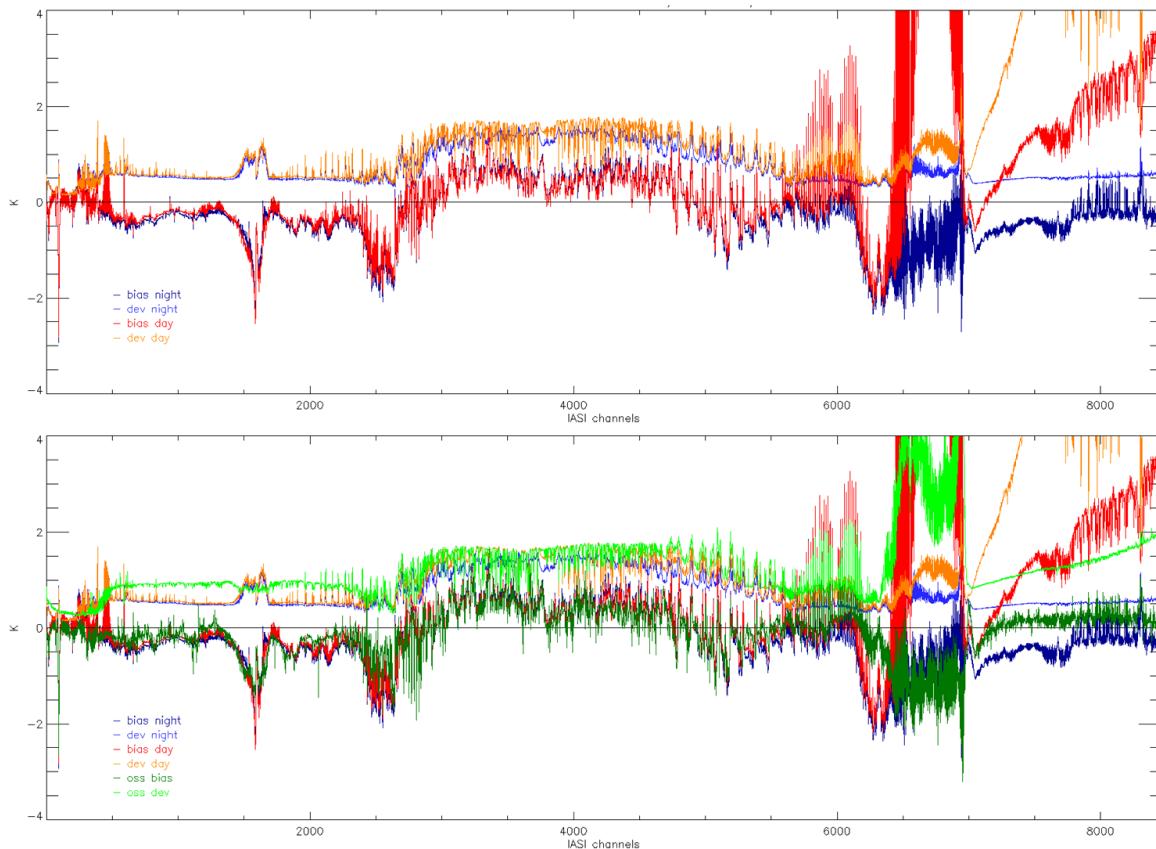
We focus hereafter on the observation error term  $y - F(x)$ , weighted by the observation error matrix  $R^{-1}$ ; where  $y$  is the observed spectrum,  $F(x)$  the forward model and  $x$  the estimated atmospheric state vector.

Because of the instrument radiometric and spectral calibrations but also of possible instrument artefacts, the spectrum recorded by the instrument may differ from the actual radiances at the top of the atmosphere. Similarly, the radiative transfer model used to represent or model these radiances with the estimated atmospheric state vector may differ from the reality because of the uncertainties in its underlying spectroscopy or errors in the radiative transfer calculations, especially in the context of a fast radiative transfer model [Mericardi 2017]. The synthetic radiances may as well be affected by the geophysical parameters not taken into account or not adjusted in the simulations (other atmospheric constituents, residual clouds, inaccurate surface parameter description...).

As a result, the observations and the simulations generally differ to a certain degree which includes a random and a static component. The random component determines within how much an observation can be meaningfully fitted by a simulation when adjusting or retrieving the true atmospheric state vector. It is captured in the matrix  $R$ . The systematic component (or bias) reflects the systematic differences between the observing and simulating systems which will remain even if the true atmospheric state vector is ingested by the forward model. This has been observed with the current IR missions but also long experienced with other instruments [SWG paper EUM/STG-SWG/35/13/DOC/10 on the radiance bias correction at EUMETSAT] and accounted for in the retrieval and assimilation community ([Eyre 1992], [Dee 2004]) because of the significance of these biases. We list hereafter a few reasons for these systematic differences and provide some results to quantify them.

- A. The specification for the absolute radiometric calibration is of 0.5 K for IRS, which is already significant compared to required the instrument noise.
- B. We present in the Figure 32 the statistics of OBS-CALC for IASI temperate and tropical ocean observations under clear-sky conditions as assessed by visual inspection of the corresponding AVHRR images. The atmospheric state vectors are built with ECMWF analyses and the forward computations are performed in turn with former versions of RTTOV 10 (top) and OSS (bottom, in green). Day and night time calculations are done separately with RTTOV (red and blue respectively), they are combined in the case of OSS. All simulations show systematic differences of up to several tenth of Kelvin in the CO<sub>2</sub>, window and water-vapour regions, sometimes exceeding 0.5 K, with rapid

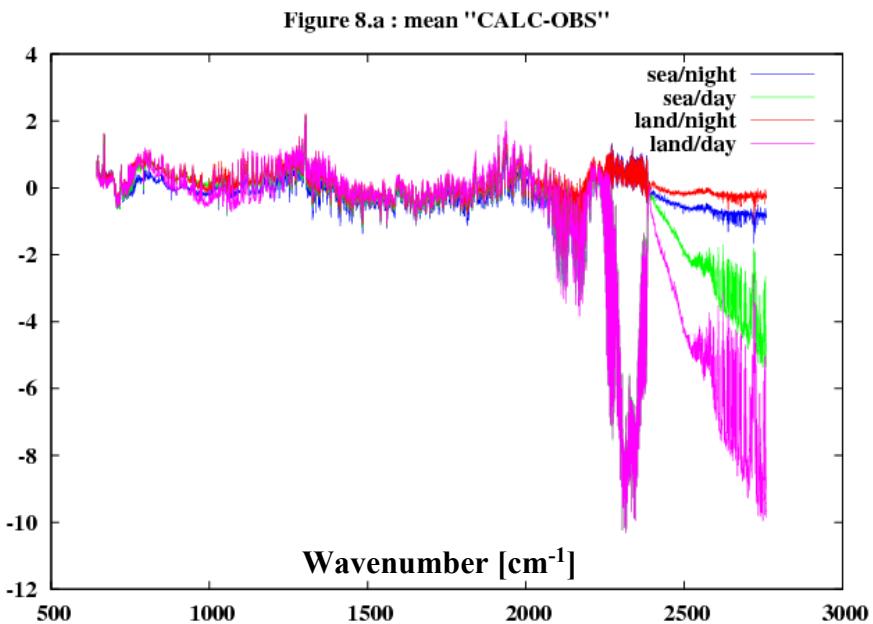
sign change in neighbouring spectral channels. The systematic differences are extreme in the O<sub>3</sub> and CH<sub>4</sub> regions, due do the inaccurate trace gas profiles used for the simulations. Handling these channels in the context of retrieval is discussed in section 3.



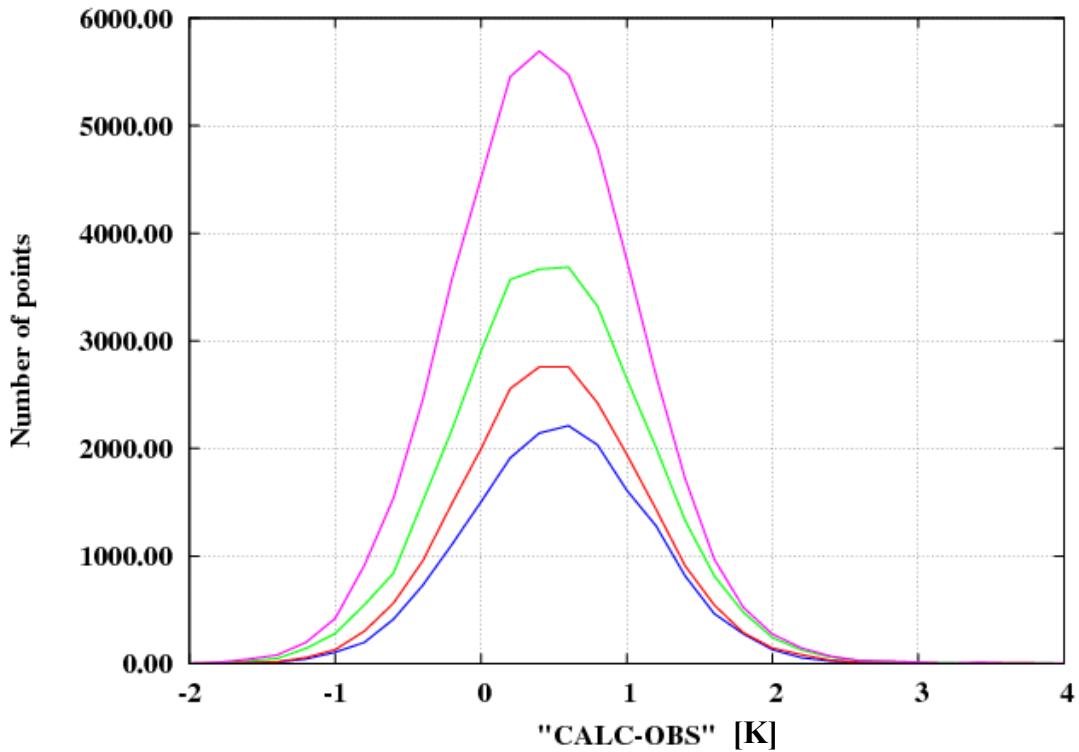
**Figure 32:** IASI OBS-CALC statistics computed with the clear ocean cases identified in the IAVISA database at mid and low latitudes.

- C. Similar results and observations are made when the atmosphere is described with sonde measurements instead of a numerical model. In this case, a multi-year extract of the ARSA database was used, which consists of quality controlled sonde measurements collocated with satellite observations at the Laboratoire de Météorologie Dynamique [Armante et al. 2013]. They are illustrated in the Figures 33 and 34 (the statistics of the CALC-OBS are represented instead of OBS-CALC). Here again, the systematic differences typically reach 0.5 K and may exceed it, with rapid sign change for neighbouring channels.
- D. In the Figure 35, we show the systematic effect on radiances of uncertainties in the forward model. The statistics (red: standard deviation, black: bias) result from the comparison of 5190 spectra simulated with LBLRTM 12.2 and 11.1 [Matricardi 2014]. It is important to note that no observations and no collocation errors are involved here. We are solely quantifying the effect of several spectroscopic upgrades, namely including the introduction of CH<sub>4</sub> line mixing, new CO<sub>2</sub> line mixing coefficients, the introduction of temperature dependence of CO<sub>2</sub> continuum in the CO<sub>2</sub> band head region, modification of CO<sub>2</sub> continuum coefficients between 2000 and 3000 cm<sup>-1</sup> and

modification of H<sub>2</sub>O self continuum coefficients between 2000 and 3200 cm<sup>-1</sup>. Here again, the systematic effects reach 0.5 K in the CO<sub>2</sub> and H<sub>2</sub>O regions, with particular stronger and severe consequences in the O<sub>3</sub> region where the biases in adjacent channels exceed 1 K.

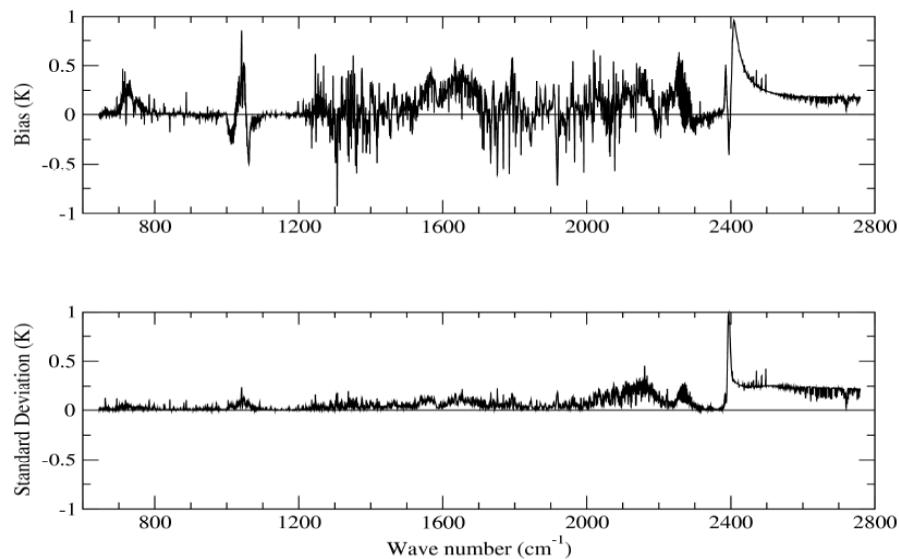


*Figure 33: Mean IASI CALC-OBS computed with ARSA sonde measurements and RTTOV*

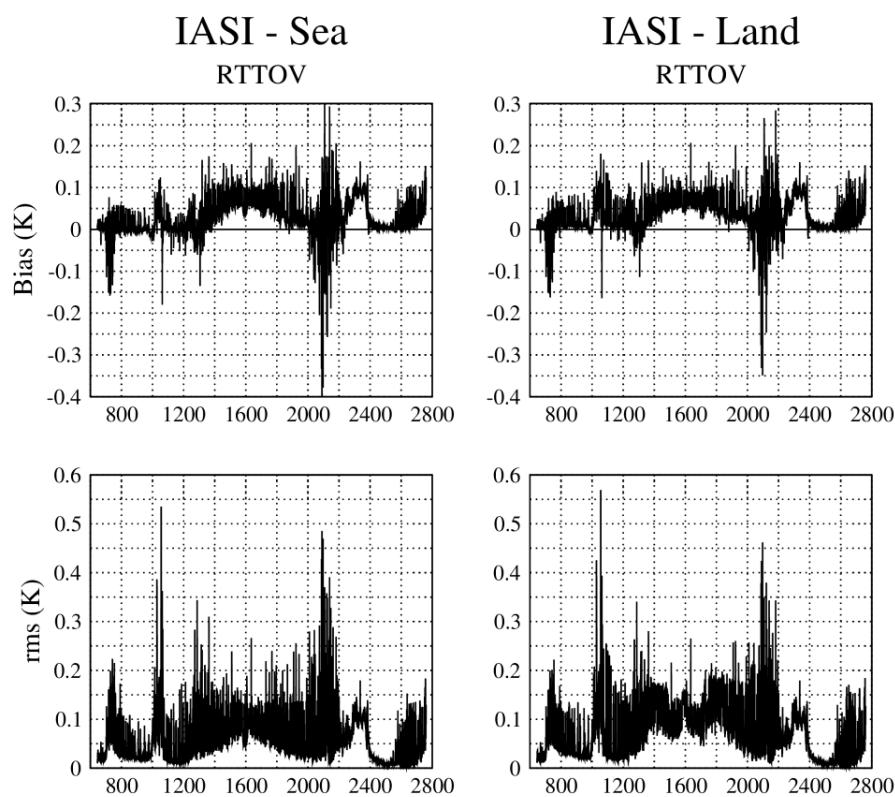


*Figure 34: Histogram of the differences IASI "CALC-OBS" at 704.5 cm<sup>-1</sup> (CO<sub>2</sub> 15 μm band) with the ARSA sonde database: green = sea day, blue = sea night, purple = land day, red = land night*

Furthermore [not shown here] it is observed that the biases between observations and simulations may vary with the viewing angle as well as with the surface and atmospheric situation. Bias correction can be implemented as a simple static offset for each channel, but more elaborated radiance tuning algorithms can be developed, for instance with a linear regression against the viewing geometry, or even aiming more sophisticated modelling for different scenes and types of observations.



**Figure 35:** Mean (top) and standard deviation (bottom) difference between LBLRTM v11.1 and v12.2 simulations with different 5190 profiles [Matricardi 2014]



**Figure 36:** Mean (top) and standard deviation (bottom) difference between RTTOV and LBLRTM v12.2 simulations of IASI maritime (left) and continental (right) spectra [Matricardi 2017]

**APPENDIX F****EXAMPLE OF A C++ IMPLEMENTATION OF THE SUB-PROFILE POINTS COMPUTATIONS**

The following pages contain listings of the C++ implementation of the algorithms defined in §4.16.

```
geoPos ongroundShift(geoPos target, double targToSatAz, double referenceAltitude, double zenithAngle)
{
// input parameters:
// =====
// target target geocentric longitude and latitude in degrees (EF)
// targToSatAz target to Satellite Azimuth in degrees (Topocentric, az counted from local north)
// referenceAltitude altitude in meters of new target
// zenithAngle target zenith angle in degrees
//
// The following data structure is used:
// geoPos:
// =====
//struct geoPos {
// double ra,de;
// };
//
//functions called:
//=====
// destVicenty:
geoPos newTarget;
double distance;
double theta;
double PI =3.141592653589793238463;
// theta: 90 deg - target zenith angle
theta = 90.0 - zenithAngle;
// distance from target to new target defined by reference altitude
distance = referenceAltitude/tan(theta*PI/180.0);
newTarget = destVicenty(target.de, target.ra, targToSatAz, distance) ;
newTarget.dist = distance;
return newTarget;
};
```

```

geoPos destVincenty(double lat1, double lon1, double bearing, double dist)
{
/*****
 * Calculate WGS 84 destination given starting lat/long (degrees),
 * bearing (degrees) & distance (Meters).
 *
 * from: Vincenty direct formula - T Vincenty, "Direct and Inverse
 * Solutions of Geodesics on the Ellipsoid with application of
 * nested equations", Survey Review, vol XXII no 176, 1975
 * http://www.ngs.noaa.gov/PUBS_LIB/inverse.pdf
 *
 *****/
/* local variable definitions */
// Trigonometric constants and conversion factors
// -----
double PI =3.141592653589793238463;
double toDEG, toRAD;
toDEG = 180.0/PI;
toRAD = 1.0/toDEG;
// WGS-84 ellipsoid
double a=6378137.0, b=6356752.3142, f=1/298.257223563;
double alpha1,sinAlpha, sinAlpha1, cosAlpha1, cosSqAlpha;
double sigma, sigma1, cos2SigmaM, sinSigma, cosSigma, deltaSigma, sigmaP;
double tanU1, cosU1, sinU1, uSq;
double A, B, C, L, lambda;
double tmp, lat2;
//double revAz; /* unused but retained for alg completeness */
double lat2out, lon2out;
geoPos newTarg;
/* code body */
alpha1 = bearing*toRAD;
sinAlpha1 = sin(alpha1);
cosAlpha1 = cos(alpha1);
tanU1 = (1-f) * tan(lat1*toRAD);
cosU1 = 1 / sqrt((1 + tanU1*tanU1));
sinU1 = tanU1*cosU1;
sigma1 = atan2(tanU1, cosAlpha1);
sinAlpha = cosU1 * sinAlpha1;
cosSqAlpha = 1 - sinAlpha*sinAlpha;
uSq = cosSqAlpha * (a*a - b*b) / (b*b);
A = 1 + uSq/16384*(4096+uSq*(-768+uSq*(320-175*uSq)));
B = uSq/1024 * (256+uSq*(-128+uSq*(74-47*uSq)));
sigma = dist / (b*A);
sigmaP = 2*PI;
while (fabs(sigma-sigmaP) > 1e-12) {
cos2SigmaM = cos(2*sigma1 + sigma);
sinSigma = sin(sigma);
cosSigma = cos(sigma);
deltaSigma = B*sinSigma*(cos2SigmaM+B/4*(cosSigma*(-1+2*cos2SigmaM*cos2SigmaM)-B/6*cos2SigmaM*(-
3+4*sinSigma*sinSigma)*(-3
+4*cos2SigmaM*cos2SigmaM)));
sigmaP = sigma;
sigma = dist / (b*A) + deltaSigma;
}
tmp = sinU1*sinSigma - cosU1*cosSigma*cosAlpha1;
lat2 = atan2(sinU1*cosSigma + cosU1*sinSigma*cosAlpha1,

```

```
(1-f)*sqrt(sinAlpha*sinAlpha + tmp*tmp));
lambda = atan2(sinSigma*sinAlpha1,
cosU1*cosSigma - sinU1*sinSigma*cosAlpha1);
C = f/16*cosSqAlpha*(4+f*(4-3*cosSqAlpha));
L = lambda - (1-C)*f*sinAlpha*(sigma+C*sinSigma*(cos2SigmaM+C*cosSigma*(-1+2*cos2SigmaM*cos2SigmaM)));
// final bearing
// revAz = atan2(sinAlpha, -tmp);
// return new position in degrees
newTarg.de = lat2*toDEG;
newTarg.ra = lon1+(L*toDEG);
return newTarg;
}
```

<Compose your document using the styles provided here for headings, captions and appendices. Ensure the last chapter of your document contains a section for implementation documents if they exist – an example is provided on the next page.>