DUE GLOBVAPOUR

Algorithm Theoretical Basis Document
L3 Merged IASI + SEVIRI

Issue 1 Revision 0

21 March 2011

Project nr: ESRIN/AO/1-6090/09/I-OL

Project Coordinator: Marc Schröder
Deutscher Wetterdienst
marc.schroeder@dwd.de

Technical Officer: Bojan Bojkov
ESA
bojan.bojkov@esa.int

Document Change Record

Document, Version	Date	Changes	Originator
DOC, v1.0	2010.11.01	Original version	Martin Stengel, Nadine Schneider, Marc Schröder

Table of Contents

1 Introduction

1.1 Purpose

This document provides the Algorithm Theoretical Basis of the Level 2 and Level 3 IASI+SEVIRI water vapour profile product for the ESA DUE GlobVapour project. Some text passages were adopted from the below listed documents, e.g. NWC-SAF User Guide, the ATBD of water vapour products of CM SAF, WACMOS Design Definition V1, the Dissertation of Schwärz (2004).

1.2 Definitions, acronyms and abbreviations

AAPP	ATOVS and AVHRR Pre processing Package
BT	Brightness Temperature
FCDR	Fundamental Climate Data Record
FOR	Field of Regard
FOV	Field of View
IASI	Infrared Atmospheric Sounding Interferometer
IC	Information content
L1	Level 1
L2	Level 2
L3	Level 3
MS	Maximum sensitivity
MSG	Meteosat Second Generation
MWR	Microwave Radiometer
NWC SAF	Satellite Application Facility on support to Nowcasting and Very Short-Range Forecasting
NWP	Numerical Weather Prediction
QM	Quality Monitoring
RTM	Radiative Transfer Model
RTIASI	Name of Radiative Transfer Model
RTTOV	Name of Radiative Transfer Model
SEVIRI	Spinning Enhanced Visible and InfraRed Imager radiometer
SPhR	SEVIRI Physical Retrieval
SSM/I	Special Sensor Microwave/Imager
TCWV	Total Column Water Vapour

WACMOS Water Cycle Multimission Observation Strategy

1.3 Applicable Documents

[AD-1] DUE GLOBVAPOUR Requirements Baseline Document (RBD), issue 1, revision 0, dated 16 April 2010.

[AD-2] DUE GLOBVAPOUR Technical Specification Document (TSD), issue 1, revision 0, dated 16 April 2010.

[AD-3] DUE GLOBVAPOUR Software Development Plan (SDP), issue 1, revision 0, dated 16 April 2010.

[AD-4] DUE GLOBVAPOUR Summary Report on Existing Algorithm Comparison and Validation Reports (SVR), issue 1, revision 0, dated 29 July 2010.

1.4 Reference Documents

[RD-1] Atkinson, N., 2010: IASI Principal Components in AAPP: User Manual, SAF NWP Document, Met Office.

[RD-2] Camy-Peyret, C., and J. Eyre, 1998: The IASI Science Plan, ISSWG report.

[RD-3] Derrien, M., Farki, B., Harang, L., Le Gleau, H. L., Noyalet, A., Pochic, D., and Sairouni, A., 1993: Automatic cloud detection applied to NOAA-11 / AVHRR imagery, Remote. Sens. Environ., 46, 246-267.

[RD-4] Collard, A. D., 1998: Notes on IASI Performance, UKMO Forecasting Research Technical Report No. 256.

[RD-5] Collard, A. D., and M. Matricardi (2005), Selection of a subset of IASI channels for near real time dissemination, in Proceedings of the Fourteenth International TOVS Study Conference, pp. 700 – 715, Beijing, China.

[RD-6] Derrien, M. and LeGléau, H., 2005: MSG/SEVIRI cloud mask and type from SAF NWC, Int. J. Remote. Sens., 26, 4707–4732.

[RD-7] EUMETSAT Satellite Application Facility on Nowcasting and Very Short Range Forecasting: Algorithm Theoretical Basis Document for PGE13 “SEVIRI Physical Retrieval Product” (SPhR) v1.0

[RD-8] EUMETSAT Satellite Application Facility on Nowcasting and Very Short Range Forecasting: Product User Manual for “SEVIRI Physical Retrieval Product” (SPhR), v1.0

[RD-9] Eyre, R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS. Q. J. R. Meteorol. SOC., 115, 1001-1026.

[RD-10] Lerner, J. A., E. Weisz, and G. Kirchengast, 2002: Temperature and humidity retrieval from simulated Infrared Atmospheric Sounding Interferometer (IASI) measurements, Journal of Geophysical Research, 107, 10.1029/2001JD900,254, 2002.

[RD-11] Li, J., 2007: Recommendation on Physical Retrieval Algorithm for SEVIRI Nowcasting Product, INM NWC SAF Visiting Scientist Activity (VSA)

[RD-12] Li, J., and H.-L. Huang, 1999: Retrieval of atmospheric profiles from satellite sounder measurements by use of the discrepancy principle, Appl. Optics, Vol. 38, No. 6, 916-923.

[RD-13] Lindenbergh, R., M. Keshin, H. van der Marel and R. Hanssen, 2008: High resolution spatio-temporal water vapor mapping using GPS and MERIS observations. International Journal of Remote Sensing, 29(8), 2393-2409.

[RD-14] Lindau, R. and M. Schröder, 2010: Objective analysis (Kriging) for water vapour products. Algorithm Theoretical Basis Document, Reference Number: SAF/CM/DWD/ATBD/KRIGING, Issue 1.0, 28 January 2010.

[RD-15] Liu, X., T. S. Zaccheo, and J.-L. Moncet, 2000: Comparison of Different Non-Linear Inversion Methods for Retrieval of Atmospheric Profiles, in Proceedings of the 10th Conference of Satellite Meteorology, pp. 293-295, Long Beach, California.

[RD-16] Matricardi, M. and R. Saunders, 1999: A Fast Radiative Transfer Model for Simulation of Infrared Atmospheric Sounding Interferometer Radiances, J. Appl. Optics, 38, 5679-5691, 1999.

[RD-17] Mueller, J. 2010: MSG Level 1.5 Image Data Format Description, Eumetsat, http://www.eumetsat.int/groups/ops/documents/document/pdf_ten_05105_msg_img_data.pdf

[RD-18] PGE13 ATBD, 2010: EUMETSAT Satellite Application Facility on Nowcasting and Very Short Range Forecasting: Algorithm Theoretical Basis Document for PGE13 “SEVIRI Physical Retrieval Product” (SPhR) v1.0.

[RD-19] Pougatchev, N. et al., 2009: IASI temperature and water vapour retrievals - error assessment and validation. Atmos. CHem. Phys. Discuss., 9, 7972 – 7989.

[RD-20] Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: Numerical Recipes in FORTRAN: The Art of Scientific Computing, Cambridge University Press, Cambridge.

[RD-21] Rodgers, C. D., 1996: Information Content and Optimization of High Spectral Resolution Measurements, in Proceedings of SPIE Conference 2830, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II, pp. 136 – 147, Denver.

[RD-22] Rodgers, C.D., 2000: Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific Publishing Company.

[RD-23] Saunders, R., Matricardi, M., and Geer, A., 2008: RTTOV 9.1 Users Guide, NWP SAF report, Met. Office, 57 pp.

[RD-24] Schmetz, J., W. P. Menzel, C. Velden, X. Wu, L. van de Berg, S. Nieman, C. Hayden, K. Holmlund, and C. Geijo, 1995: Monthly mean large scale analysis of upper troposphere humidity and wind field divergence derived from three geostationary satellites, Bull. Amer. Meteor. Soc., 76, 1578 – 1584.

[RD-25] Timmermans, J. et al. 2010: Support to Science Element Water Cycle Multi-mission Observation Strategy (WACMOS), Design Definition V1, 15.01.2010.

[RD-26] Schluessel P., X Calbet, T. Hultberg, A. Arriaga, T. August, O. Oduleye, and H. Murata, 2008: EPS Product Validation Report: IASI L2 PPF, Doc.No. : EUM/MET/REP/07/0224, Issue : v2C, 5 August 2008, available at: www.eumetsat.int

[RD-27] Schwärz, M., 2004: Joint Temperature, Humidity, Ozone, and Sea Surface Temperature Retrieval from Infrared Atmospheric Sounding Interferometer Data, Dissertation, University of Graz.

[RD-28] Spencer, R. W., and W. D. Braswell, 1997: How dry is the tropical free troposphere? Implications for global warming theory, Bull. Amer. Meteor.Soc., 78, 1097 – 1106.

[RD-29] Validation Report, 2009: Validation Report – Vertically Integrated water vapour from SSM/I, EUMETSAT-CMSAF, Reference Number SAF/CM/VAL/HTW_SSMI_global_DS, Issue 1.1, 6 January 2009.

[RD-30] WACMOS DDF, 2010: Support to Science Element Water Cycle Multi-mission Observation Strategy: Design Definition. Reference Number: WACMOS_DD, issue 2.31.

[RD-31] Weisz, E., 2001: Temperature Profiling by the Infrared Atmospheric Sounding Interferometer (IASI): Advanced Retrieval Algorithm and Performance Analysis (ph.d. thesis), Tech. Rep. Wissenschaftl. Ber. No. 11, Institute for Geophysics Astrophysics, and Meteorology, Institute of Physics, University of Graz, Austria.

[RD-32] Weisz, E., G. Kirchengast, and J. A. Lerner, 2003: An efficient channel selection method for Infrared Atmospheric Sounding Interferometer data and characteristics of retrieved temperature profiles, Tech. Rep. Wissenschaftl. Ber. No. 15, IGAM, University of Graz, Austria.

[RD-33] Whyte, K. and N. Atkinson, 2006: AAPP Overview, SAF NWP Document, Met Office.

1.5 Structure of the document

Section 2 gives a short overview of algorithms used, illustrated with the processing flow. The algorithm description occurs in section 3, where all theoretical and practical important steps of the algorithm of both instruments will be explained in detail. Assumptions and limitations can be found in section 4. The conclusion is written in section 5.

2 Algorithm overview

The vertically resolved water vapour profiles of the instruments SEVIRI as well as IASI are retrieved through a 1D-Var algorithms (see flow chart in Figure 1). Using the NWC-SAF retrieval scheme for SEVIRI and EUMETSAT IASI Level 2 products a merging technique following Lindenberg et al. (2008) and Lindau and Schröder (2010) was developed within the ESA STSE WACMOS project (see WACMOS DDF, 2010 for details).

Within the GlobVapour project, an IASI assessment has been executed. Five different IASI water vapour profile algorithms developed by the project contractors and outside groups have been validated and compared. The selected candidate of the IASI assessment is the algorithm developed by Marc Schwärz (U. Graz) which replaces the EUMETSAT IASI Level 2 product within the merged SEVIRI+IASI product.

The Level 2 data set of both instruments are gridded and vertically integrated to provide swath-based information of water vapour for three vertical layers (surf. - 850 hPa, 850 hPa - 500 hPa, 500 hPa - 200 hPa) and the upper and lower layer boundaries, including additional information.

Furthermore, a bias correction is performed to support an optimal Kriging performance, also described in described in WACMOS DDF (2010).

The merged L3 end product of IASI and SEVIRI undertakes grid-based global data with updated GlobVapour metadata information in netCDF format following the CF-1.4 standard for three-hourly as well as monthly means. Each file provides four variables: TCWV, TCWV error, quality flag , and Number of Observation.

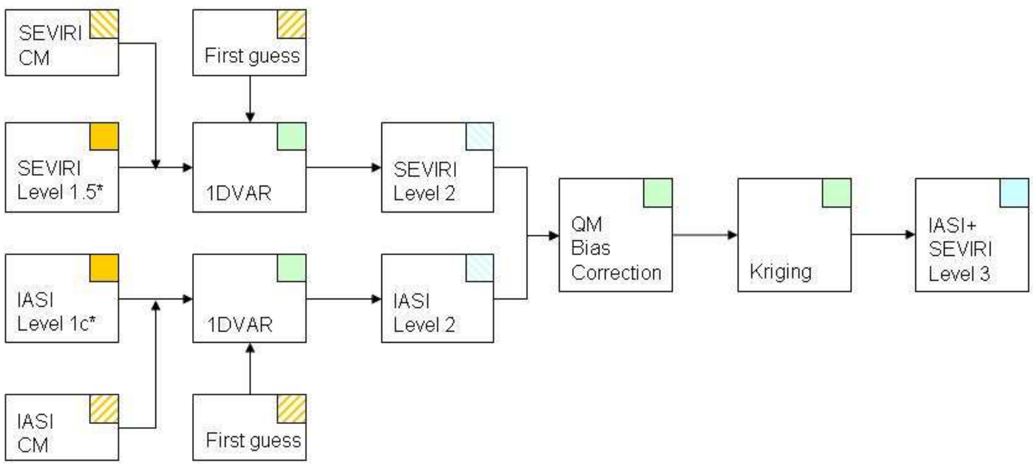

Figure 1: Flow chart for the merged IASI-SEVIRI profile product. Input data is marked orange, products are marked blue (blue shaded: instantaneous level 2 products) and software development is marked green. Orange shading marks higher level input data (QM: quality monitoring, CM: cloud mask).

3 Algorithm description

3.1 Theoretical description

SEVIRI (Spinning Enhanced Visible and InfraRed Imager) radiometer

SEVIRI radiometer, onboard MSG, was primarily designed to support operational meteorology applications. It has 3 visible/near infrared and 8 thermal infrared channels with a spatial resolution at nadir of 3km and one visual broadband channel (HRV) where the spatial resolution at nadir is 1km. The nominal coverage is shown in Figure 2. Data is available in near real time every 15 minutes and operationally processed to level 1.5, i.e. corrected for radiometric and geometric non-linearity (Müller, 2010).

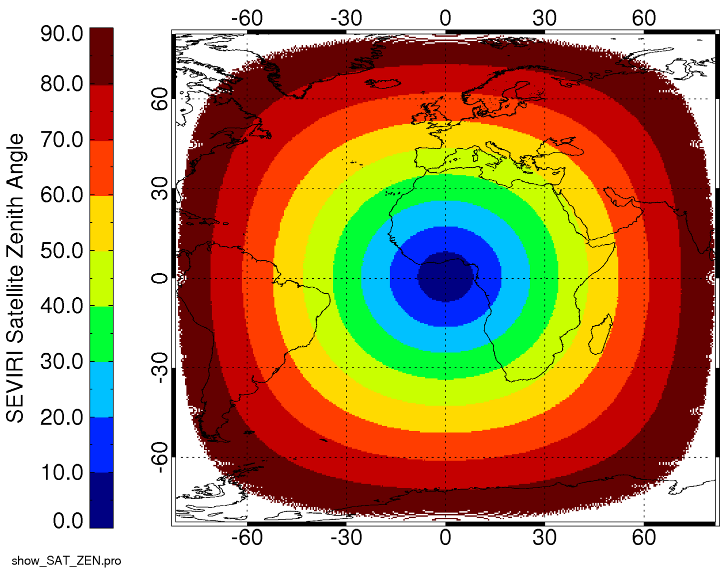

Figure 2: SEVIRI satellite zenith angles (from Timmermans et al., 2010).

IASI (Infrared Atmospheric Sounding Interferometer)

IASI, onboard MetOP, observes the atmosphere in the infrared (3.7 - 15.5 µm) in 8461 channels globally. Currently, it is the most accurate infrared sounding interferometer in orbit and allows to measure the atmosphere temperature within 1° K and relative humidity within 10 % for each slice of 1 km height.

One of the primary objectives of the IASI instrument, according to the IASI science plan (Camy-Peyret and Eyre, 1998), is the improvement of the vertical resolution of temperature and water vapor profiles to about 1 km in the middle and lower troposphere as well as improving the retrieval accuracy to within 1 K in temperature and 10 % in specific humidity. The scientific motivation of this is based on the key role of water vapor in the upper troposphere and its effects on the global climate since only small changes in humidity and its trends have serious implications on the amount of thermal energy escaping to space (Schmetz et al., 1995, and Spencer and Braswell, 1997). Additionally, this level of performance will greatly assist numerical weather prediction (NWP) in delivering accurate and frequent temperature and humidity profiles for operational and research needs and it will supply more accurate quantifications of the climate variability, particularly contributing to our knowledge of the climate of the upper troposphere.

SEVIRI 1D-Var retrieval scheme

This section largely follows WACMOS DDF (2010) and PGE13 ATBD (2010).

The SEVIRI 1D-Var module is based on an existing 1D-Var retrieval scheme (SEVIRI Physical Retrieval (SPhR)) developed in the framework of the NWC SAF. The core of the SPhR was developed for the WACMOS project by the Satellite Application Facility on support to Nowcasting and Very Short-Range Forecasting (NWC SAF), and a beta version of the algorithm was distributed as Product Generator Element (PGE) 13 of the NWC SAF MSG software package to users in summer 2009. A complete description of the retrieval algorithm can be found in the Algorithm Theoretical Basis Document for PGE13 of EUMETSAT.

The 1D-VAR technique uses the variational principle to solve the retrieval problem (Eyre, 1989) and is applied to SEVIRI observations. Based on a first-guess profile, which is a 6-hour forecast from a NWP model, the 1D-VAR method determines the most probable temperature and humidity profiles.

A short summary, based on the PGE13 ATBD, is given in the following:

The input parameters for the PGE13 SPhR are the SEVIRI brightness temperatures from six infrared channels centred at 6.2, 7.3, 8.7, 10.8, 12.0 and 13.4 μm. If one or more channels are missing, no results are produced. Numerical weather prediction (NWP) or reanalysis model data consisting of surface pressure and skin temperature as well as the temperature and relative humidity at various vertical levels need to be provided as additional input. The land-sea mask (Global 30 Arc-Second Elevation GTOPO30) and emissivity maps for each month and each IR channel (Seemann et al., 2008) that are also needed are available within the NWC SAF software package. Also mandatory is the provision of a cloud mask as the water vapour retrieval can only be performed under clear sky conditions. Within the NWC SAF MSG software package, the cloud mask is derived before the water vapour module is executed. The WACMOS software utilises the PGE13 package only and relies on the already existing SEVIRI cloud mask from the CM SAF.

The PGE 13 SPhR is a physical algorithm based on optimum estimation (OE) theory. This inversion technique iteratively finds the atmospheric temperature and moisture profile that best reproduces the IR observations by using a radiative transfer model. Because the retrieval problem is ill-posed, additional information is needed to constrain the solution. Often, this is accomplished by means of a background profile obtained from a climate mean, a regression technique, and/or numerical forecast products. It further requires a first-guess profile to initiate the iteration process. A flow chart of the SEVIRI Physical Retrieval is displayed in Figure 3.

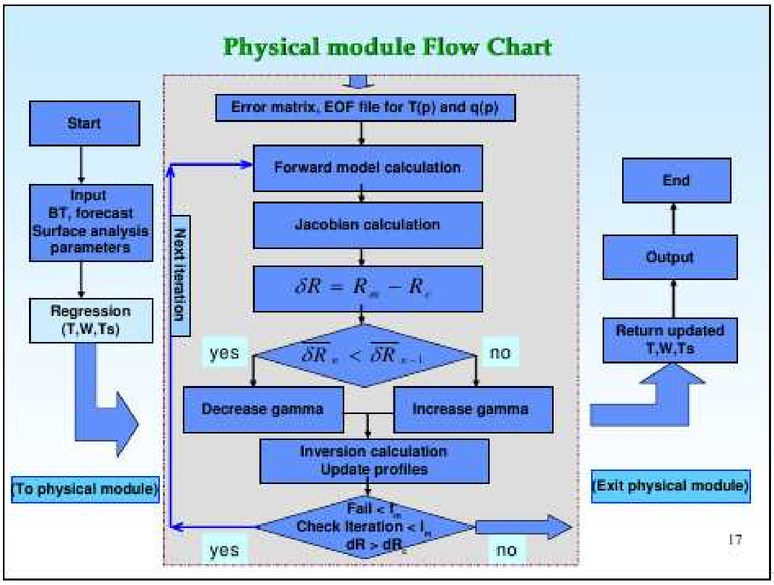

Figure 3: Flowchart of the SEVIRI Physical Retrieval taken from Li (2007).

Basically, the PGE 13 retrieval involves two steps:

- A non-linear regression is performed to build a first-guess water vapour profile.
- An iterative physical solution of the radiative transfer equation is accomplished using the results of the first step as initial profiles.

A typical first-guess field is a short-term NWP forecast, however, it was found that a regression combining forecast and SEVIRI IR BTs is usually better than the forecast itself. Since SEVIRI has a few sounding spectral bands, the temperature/moisture profiles from NWP forecast model are used as additional predictors. Here, the temperature forecast between 100 and 1013 hPa and the specific humidity forecast between 300 and 1013 hPa are used as additional predictors. Given *Z* (e.g., temperature or mixing ratio at a given pressure level) as a predictand, the regression equation has the following format:

$$Z = A_0 + \sum_{j=1}^{N} B_j T_{bj} + \sum_{j=1}^{N} C_j T_{bj}^2 + \sum_{l=1}^{ntemp} b_{tl} T_l + \sum_{l=1}^{nq} b_{wl} \log(q_l) + D_1 p_s + D_2 \sec\theta + D_3 \cos(Lat) \quad (3.1)$$

where T_b is the brightness temperature; T and q are forecast temperature and specific humidity respectively. p_s, θ and Lat are the surface pressure, local zenith angle and latitude, respectively. *A*, *B*, *b*, *C* and D_i are regression coefficients which are determined using a global radiosonde dataset (Li, 2007). *N*, *ntemp* and *nq* are the total number of SEVIRI IR spectral bands, the number of temperatures and the number of specific humidity information, respectively.

Following the NWC SAF ATBD, the first-guess is generated with the regression technique and then a nonlinear iterative physical retrieval is performed. Given the vector of measured SEVIRI radiances, $\boldsymbol{Y}^m$, and the observation error covariance matrix $\boldsymbol{E}$ (which is the sum of the instrument noise and the forward model error) as well as a background of the atmospheric profile, $\boldsymbol{X}^b$ together with the background error covariance matrix $\boldsymbol{B}$, an approximate estimate of the true atmospheric profile, $\boldsymbol{X}$, is obtained by minimizing the cost function J($\boldsymbol{X}$) (Rodgers, 2000)

$$J(X) = \left[Y^m - F(X)\right]^T E^{-1}\left[Y^m - F(X)\right] + \left[X - X^b\right]^T \gamma B^{-1}\left[X - X^b\right] \quad (3.2)$$

where the superscript *T* denotes the transpose of a matrix. As forward model F, the radiative transfer model RTTOV version 9.3 (Saunders et al., 2008) maintained by the NWP SAF is used.

Using a Newtonian iteration the state vector X is found by applying:

$$\delta X_{n+1} = \left(F_n^{'T} \cdot E^{-1} \cdot F_n^{'} + \gamma B^{-1}\right)^{-1} \cdot F_n^{'T} \cdot E^{-1} \cdot \left(\delta Y_n + F_n^{'} \cdot \delta X_n\right) \quad (3.3)$$

where n denotes the iteration step, i.e., n=0 is the first-guess profile. The definitions

$$\begin{aligned} \delta X_n &= X_n - X^b; \\ \delta Y^m &= Y^m - F(X_n) \end{aligned} \quad (3.4)$$

were introduced as well. $\boldsymbol{F'}$ is the tangent linear operative (Jacobian) of the forward model $\boldsymbol{F}$. Rearranging Equation (3.3), as described in the PGE13 ATBD, making use of correlations among the atmospheric variables, results in an expression that allows the determination of $\boldsymbol{X}$ with a minimum of matrix multiplications and computation time. The regularization parameter γ is introduced to speed up the convergence and increase the numerical stability of the inversion process. It is adjusted in each iteration according to the discrepancy principle (Li and Huang, 1999). Finally, the error covariance matrix, $\boldsymbol{S}$, of the estimate $\boldsymbol{X}$ is determined (Rodgers, 2000)

$$S = \left(F^{T} \cdot E^{-1} \cdot F + B^{-1}\right)^{-1} \qquad (3.5)$$

S finally provides the expected retrieval error for the temperature and humidity profile that can be used in the objective analysis to provide errors on the final product grid.

A major processing step is the generation of the SEVIRI Level 2 water vapour profiles using the new NWC SAF SEVIRI Physical Retrieval. SEVIRI carries out observations with a spatial resolution of 3x3 km at nadir every 15 minutes. NWC SAF PGE13 products can be generated at every single time step on a Field-of-Regard (FOR) basis. One FOR consists of M x M SEVIRI field of views (FOVs). The FOR size is configurable to the user requirements. The default FOR size is 3 by 3 FOVs which is equivalent to approximately 9x9 km spatial resolution. The mean of all clear pixel brightness temperatures (BTs) within each FOR (default and used here) or the BT of the warmest clear sky pixel of channel IR10.8 are the two available methods to determine the BT of an FOR. Within the scope of WACMOS, the SEVIRI Level 2 data is generated every three hours. A better temporal resolution is in principle possible and might be considered in a later stage. The default FOR of 3 by 3 SEVIRI FOVs has been considered as adequate (PGE13 ATBD) and is kept. This is a reasonable choice for the merged SEVIRI+IASI product, which should cover the whole MSG disc with a spatial resolution of $(0.25^{\circ})^2$. The size of the FOR has a large influence on the computation time. Assuming a FOR of 3x3 FOVs instead of processing every single SEVIRI pixel reduces the computation time of the MSG PGE13 SPhR from approximately three hours to 30 minutes on a single processor.

Besides the size of the FOR and the determination of the BTs of each FOR, a number of additional parameters, summarized in a configuration file, can be customized to the users' needs. E.g. a threshold error between the calculated RTTOV brightness temperatures and the observed SEVIRI BTs (BT_RMS_THRESHOLD keyword) has to be defined. Only if the difference between the calculated BTs from the first-guess profiles and the SEVIRI BTs are greater than this threshold, the physical retrieval is performed. The maximum number of iterations and a residual threshold (MAX_RESIDUAL keyword) are also configurable parameters in the PGE13 configuration file. Using the default values of the *BT_RMS_THRESHOLD* and *MAX_RESIDUAL* keyword implicates that no physical retrieval is performed for a large number of clear sky pixel. Instead, the result of the first-guess is written as final output (as indicated by the blue points on the MSG disc displayed on the left hand side of Figure 4. Since the error of the SEVIRI water vapour profile, which is essential for the Kriging, is only obtained when the physical retrieval is performed, the values of *BT_RMS_THRESHOLD* was reduced to 0.001. As a consequence the physical retrieval is applied at all clear sky pixels (see Figure 4, right picture).

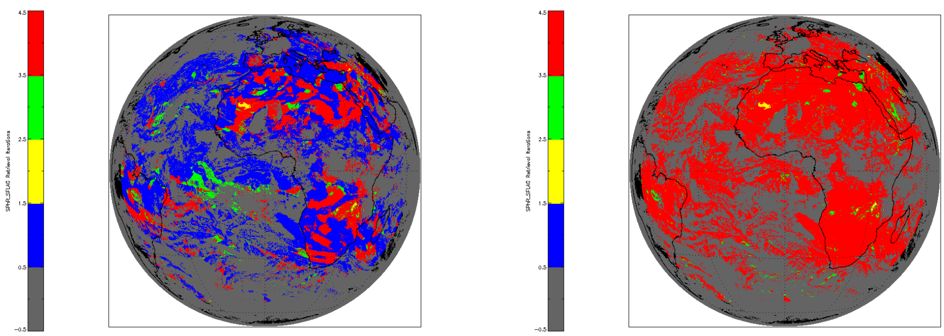

Figure 4: Parameter retrieval iterations contained in the flag SPhR_SFLAG for the SEVIRI slot at 11h45 UTC on August 01 2008. Pixels processed with the first-guess are blue. Pixels processed with the physical retrieval are shown in yellow (1 iteration), green (2 iterations) and red (3 iterations). Left: Result with the default values of BT_RMS_THRESHOLD (2.0) and MAX_RESIDUAL (1.0). Right: Results with a BT_RMS_THRESHOLD and a MAX_RESIDUAL value of 0.001. In both cases a FOR of 3 x 3 SEVIRI FOVs was assumed.

The main PGE13 SPhR output is written to hdf5 files including e.g. the total column water vapour from the retrieved profiles. Besides, binary files can be generated. Here, the information from three different binary files, described in Table 1 are used. Further details on the file format, etc. can be found in the PGE 13 SPhR Product User Manual.

Table 1: Description of the binary output files obtained from the PGE13 SPhR. The parameters given in bold, namely the surface pressure, the specific humidity (q) profile and the specific humidity error profile are used within GlobVapour. A more detailed description of the files can be found in the Product User Manual of PGE 13 SPhR.

Name of binary file	Description
PGE13_CLEAR_FOR_background	Background NWP T and q (ppmv) profiles at 43 RTTOV levels + Ozone (43) profile (empty) + 2m temperature + 2m q + **surface pressure** + skin temperature
PGE13_CLEAR_FOR_end	Final T (K) and **q (ppmv) profiles** and skin temperature.
PGE13_CLEAR_FOR_error_estim	Final T (K) and **q (ppmv) error profiles** and skin temperature error.

Detailed information can be found in the Design Definition V1 of WACMOS (Timmermans, 2010).

IASI 1D-Var retrieval scheme

The chosen IASI retrieval scheme is developed by Marc Schwärz (Schwärz, 2004) which is the reference document for this section.

The basics of this 1D-Var method are similar to the algorithm design of the 1D-Var for the instrument SEVIRI as described above.

The IASI retrieval involves three main steps:

- The AAPP pre-processing for generating the input data for the 1D-Var (Level 1d).
- A collocation of background data for determining the first guess information.
- An iterative physical retrieval for calculating the Level 2 product.

It should be mentioned, that the formally implemented transmittance model RTIASI was replaced by the RTM RTTOV (detailed information can be found in the publication of Saunders, 1999). The specific main technical alignments of the 1D-var for IASI will be described in the following. Some text passages are along the lines of Schwärz (2004):

A physical-statistical approach has been utilized in the retrieval, which is incorporating the underlying physics and modeling the uncertainties in the measurements as well as in the prior knowledge of the state. Preceding this inversion, a channel selection scheme was a implemented to reduce the number of IASI channels used, for removing redundant information for performance and computational reasons.

<u>Variational approach used</u>

With the assumption of a linear problem illustrating by the Bayesian approach

$$\mathrm{y} = \mathrm{K}\mathrm{x} + \epsilon \tag{3.6}$$

where x is the state of the physical system, and ϵ is defined as the measurement noise, the vectors and matrices can be set up illustrated in the following, where y depends only on one set of parameters, e.g. temperature profile, x_T , where T refers to temperature

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} (& \mathrm{k}_{T,1} &) \\ & \vdots & \\ (& \mathrm{k}_{T,m} &) \end{pmatrix} \begin{pmatrix} x_{T,1} \\ \vdots \\ x_{T,n} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix} \tag{3.7}$$

with

$$\mathrm{k}_{T,i} = \begin{pmatrix} \frac{\partial y_i}{\partial x_{T,1}} & \cdots & \frac{\partial y_i}{\partial x_{T,n}} \end{pmatrix} \tag{3.8}$$

the so-called weighting functions, which give the dependence of one measurement y_i, on the atmospheric profile x_T.

Since a moderately non-linear problem has to be solved in this case, the used inversion method is an iterative optimal estimation algorithm (Gauss-Newton method). The iterative progression is formulated as:

$$\mathbf{x}_{i+1} = \mathbf{x}_{ap} + \mathbf{S}_i\, \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \left[(\mathbf{y} - \mathbf{y}_i) - \mathbf{K}_i \left(\mathbf{x}_{ap} - \mathbf{x}_i \right) \right] \quad (3.9)$$

with

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i) \text{ and } \mathbf{S}_i = \left(\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1} \right)^{-1} \quad (3.10)$$

Hence, the only matrix left with a special setup in the joint representation is the a priori error covariance matrix S_{ap}. It is constructed as a block-diagonal matrix based on the assumption that there are no cross correlations between temperature, humidity (used in form of specific humidity), and ozone:

$$\mathbf{S}_{ap} = \begin{pmatrix} \left(\mathbf{S}_{ap,T} \right) & 0 & 0 \\ 0 & \left(\mathbf{S}_{ap,H} \right) & 0 \\ 0 & 0 & \left(\mathbf{S}_{ap,O} \right) \end{pmatrix} \quad (3.11)$$

For the measurement error covariance matrix S_{ap}, the same setup applies as for the *a priori* error covariance matrices except that the measurement error covariance matrix is modified in its diagonal according to Liu et al. (2000)

$$\mathbf{S}_\epsilon(k,k) = \max \left[\frac{(\mathbf{y}(k) - \mathbf{y}_i(k))^2}{\alpha}, \sigma^2(k) \right] \quad (3.12)$$

where k is the channel index, i is the iteration index, σ is the measurement noise, i.e. the original diagonal element of S_ϵ, and a is a control parameter which is set to 4 in this study, following Weisz et al. (2003). This correction is implemented in the algorithm in the way that at each iteration step the difference between the actual measurements, i.e. the measurements obtained by forward modeling the state of the previous iteration, and the "true" measurements is compared with the measurement error, and then the larger value of the two quantities is taken to be the diagonal element of the measurement error covariance matrix.

Following the convergence criterion given by *Rodgers* (2000):

$$\chi^2 \leq m \quad (3.13)$$

where *m* is the number of used channels and the cost function x^2 is obtained by:

$$\chi^2 = (\mathbf{y} - \mathbf{y}_i)^T \mathbf{S}_\epsilon^{-1} (\mathbf{y} - \mathbf{y}_i) + (\mathbf{x}_i - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1} (\mathbf{x}_i - \mathbf{x}_{ap}) \tag{3.14}$$

If this criterion is not met, the iteration loop is terminated either if

$$\chi_i^2 \geq \chi_{i-1}^2 \tag{3.15}$$

or the number of iterations, *i*, exceeds a certain threshold.

Figure 5 shows the interpolated IASI level 1c noise values, their modification according to the actual brightness temperature, and these values after adding the 0.2 K forward model error for a U.S. standard mid-latitude summer atmosphere, since the temperature modified level 1c noise values are superposed with an 0.2 K forward model error to roughly account for errors in the forward model (Collard,1998, and Weisz, 2001).

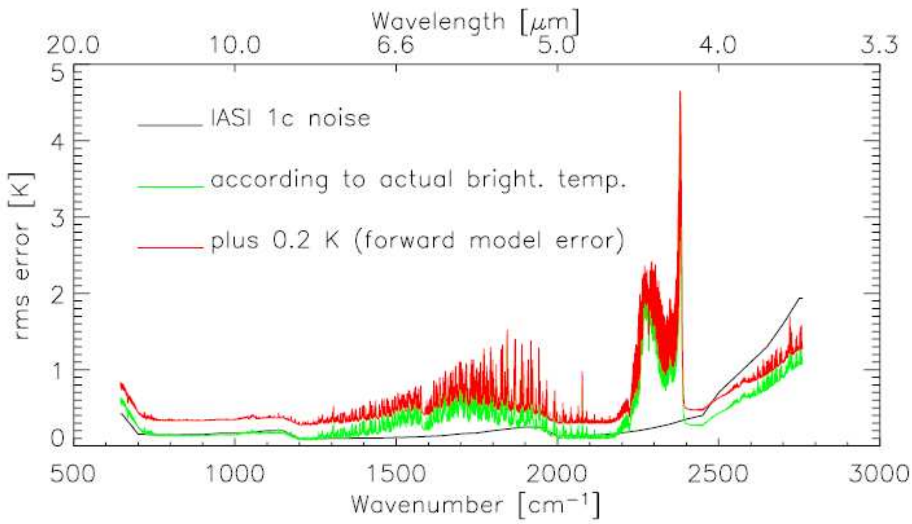

Figure 5: Interpolated IASI level 1c noise values, their modification according to the actual brightness temperature, and these values after adding the 0.2 K forward model error for a U.S. standard mid-latitude summer atmosphere (from Schwärz, 2004).

Since the full IASI spectra contain 8461 channels it is essential to reduce this number and remove redundant information for performance and computational reasons. Press et al. (1992) point out that sets of linear equations can be routinely solved for a dimension of the matrices, N, as large as a few hundred with double precision (64 bits) representation if the equations are not close to singular (i.e. some of the equations are close to linearly dependent).

Losses in performance occur particularly during the process of matrix inversion, which is a N_3 problem as well as in the gradient matrix modelling which depends also not linearly on the number of channels (the increase of the number of channels from about 300 to about 1800 causes a factor of about 12 increase in CPU time, see **Fehler! Verweisquelle konnte nicht gefunden werden.**. Note that this results also from numerical instabilities which cause an increase in iteration loops in the retrieval).

With the luxury of high spectral resolution those channels – 1220- 1370 cm^{-1} (N2O, CH4, and SO2) and 2085-2220 cm^{-1} (CO and N2O) – whose "foreign" gas emissions contribute significantly to the measured brightness temperatures (see Table 2) can also be excluded. The associated trace gas constituents can be considered as uncertainties in the temperature and humidity profile retrieval since they are only modelled approximately in the used version of RTIASI.

The channel subset used in this implementation of the retrieval scheme is based on the proposal by Collard et al. (2005). Here, the channels are spectrally thinned and filtered with respect to the forward model performance of the radiative transfer model.

Table 2: IASI Spectral Range (source: http://www.esa.int and http://smsc.cnes.fr/IASI/).

Spectral Range [cm^{-1}]	Primary Application
650 - 770	Temperature sounding (CO_2 band)
770 - 980	Surface and cloud properties
1000 - 1070	O_3 sounding
1080 - 1150	Surface and cloud properties
1210 - 1650	Water vapor and temperature sounding (and N_2O, CH_4, and SO_2)
2100 - 2150	CO column amount
2150 - 2250	Temperature sounding and N_2O column amount
2350 - 2420	Temperature sounding
2420 - 2700	Surface and cloud properties
2700 - 2760	CH4 column amount

3.2 Practical application

L1 data source

SEVIRI and IASI Level 1.5 data were used from the DWD data archive.

The ATOVS and AVHRR Pre-processing Package (AAPP) has been used to pre-process selected IASI Level 1c radiances. The IASI processing is possible since AAPP version 6.1, released in October 2006. Thus, the data will be decoded from BUFR and stored in PC binary format using a set of reference eigenvectors to complete the output into a full IASI spatial sampling with 120 spots per scan.

IASI L1 bias correction is performed using pre-calculated correction coefficients which were inferred from relating the brightness temperature observations minus simulations with certain atmospheric properties, the so-called predictors, e.g. surface temperature, mean tropospheric temperature. Trough a regression scheme the coefficients are then calculated. With this bias correction approach systematic shortcoming of the radiative transfer model can be identified an later, when applying the bias correction scheme, significantly reduced.

Cloud mask (CM)

The SEVIRI cloud mask is available from the CM SAF. The cloud mask algorithm, implemented in the MSG package developed cooperatively by the SAF on Support to Nowcasting and Very Short-Range Forecasting (NWC SAF), is described in Derrien and LeGléau (2005). Fundamental principles of the algorithms applied to SEVIRI raw data can already be found in an earlier paper by Derrien et al. (1993). The CM SAF cloud mask is available from the DWD archive every hour at hh:45. The reason why the slot at hh:45 was chosen is due to the scanning mechanism of SEVIRI. The images are taken from south to north and east to west within approximately 12 minutes. It is argued that the scan starting at hh:45 delivers data over Europe which are closer in time to the full hour than measurement during the scan starting at hh:00.

For IASI the EUEMTSAT L2 cloud mask is used to identify clear-sky footprints in which the retrieval is run. As a second option of the IASI retrieval system the McNally and Watts cloud detection is included. The latter option is assumed to give better cloud identifications. However, it strongly depends on an accurate bias correction.

First guess input

Background data for the 1D-Vars are provided by means of ERA-Interim (IASI) and GME (SEVIRI) forecasted atmospheric and surface fields. The data is collocated to the corresponding SEVIRI pixels and IASI footprints. Within the IASI 1D-Var a module providing the necessary data from climatological values is present.

L2 data

The L2 output of the 1D-Var for the SEVIRI and the IASI water vapour product provides disc-based and swath-based information, respectively, for each day of the instrument covered area. It contains mainly information about the total column water vapour, including additional information. The unit of water vapour is kg/m^2.

For the GlobVapour project, the IASI 1D-Var calculations are performed on a reduced resolution to decrease the computational costs.

Quality monitoring (QM) and bias correction

The SEVIRI and IASI data have undergone the QM and Bias Correction module from ESA STSE WACMOS. Some QM properties of this module are: a threshold test rejecting all retrieval with exceeding pre-defined limiters and an optional outlier check. Furthermore, a bias correction was performed to support an optimal Kriging performance. Here, the biases between the SEVIRI and IASI retrievals were investigated, while the latter serves as trusted reference. Subsequently, the determined biases of the SEVIRI product were removed.

Regarding the QM of the SEVIRI Level 2 products, several checks performed within the PGE13 SPhR can be quoted: Only if the pixel or FOR is labelled as cloud free and the satellite zenith angle is below the configurable maximum threshold (default: 70°), the PGE 13 is executed. It is also checked if the retrieved profiles are between certain limits and make physical sense. In case of the water vapour profiles each level is checked for super-saturation. To check the product quality, different flags are included in the output files. Using the binary files given in Table 1 ensures that only clear-sky pixels are used.

In case of IASI, the flags *FLG_IASIBAD*, *FLG_IASICLD*, *FLG_RESID* and *FLG_SATMAN* are evaluated in order to select only good quality, clear sky IASI vertical profiles. The other flags are ignored.

Concerning water vapour bias correction systematic differences between the two datasets to be merged need to be eliminated before entering the Kriging approach. Based on the gridded data, the monthly mean of the individual data sets is calculated for each pixel. In a second step the bias is subtracted from the SEVIRI measurements, i.e., IASI is used as reference estimate because better quality water vapour information is expected. In case of a IASI bias, SEVIRI would be used as reference.

L3 data output - end product

For the data products an objective analysis method (Kriging) was applied to achieve an interpolation with the possibility to additionally provide uncertainty information for the merged products. The approach of Lindenbergh et al. (2008) was employed who merged IASI TCWV observations with hourly ground based GPS observations. This method was combined with the currently used CM SAF operational Kriging approach described in Lindau and Schröder (2010).

The first step of the sensor merging procedure is the determination of the horizontal and temporal correlation functions. The correlation functions are determined on monthly basis by fitting exponential functions to the data. The correlation analysis is done separately for the three vertical layers. Spatiotemporal interpolation into non clear-sky regions is a challenging task, mainly due to anisotropic correlations and strong gradients in water vapour content, e.g., in the vicinity of fronts. In principle, gaps can be filled by the objective analysis. However, Sohn and Bennartz (2008), Mieruch et al. (2010), and Xavier et al. (2010) demonstrated a systematic difference between TCWV within clouds and in clear sky regions. In addition, this difference is a function of location. Therefore, gap filling of clear sky only products is not feasible.

In principle, kriging can be regarded as a prediction of a value x at a location p_0 by using information at surrounding positions p_i. Based on SEVIRI and IASI measurements, the kriging theory is explained in the following on the exemplary basis of total column water vapour. In praxis, the algorithm is applied to all three layers individually.

The integrated total water vapour column $x(p_0,t_S)$ at location p_0 and time t_S is estimated as a linear combination

$$x(p_0,t_S)=\sum_{i=1}^{n}\lambda_i\,[\,x_S(p_i,t_S)+\Delta x_S(p_i,t_S)\,]+\nu\,[\,x_M(p_0,t_M)+\Delta x_M(p_0,t_M)\,] \quad (3.16)$$

of SEVIRI observation $x_S(p_1,t_s)$, $x_S(p_2,t_s)$, …, $x_S(p_n,t_s)$ made at n different locations (p_i, i=1,..n) at time t_s and one IASI observation $x_M(p_0,t_M)$ at location p_0, obtained at time t_M. Δx_S and Δx_M denotes the SEVIRI and IASI retrieval error, respectively.

$$\sum_{t=1}^{m}\left(x(p_0,t_S)-\sum_{i=1}^{n}\lambda_i\,[\,x_S(p_i,t_S)+\Delta x_S(p_i,t_S)\,]-\nu\,[\,x_M(p_0,t_M)+\Delta x_M(p_0,t_M)\,]\right)_t^2=min \quad (3.17)$$

The weights λ_i and ν are determined by differentiating Equation (3.1) with respect to all weights. This leads to n+1 linear equations which can be written in matrix form (Equation **Fehler! Verweisquelle konnte nicht gefunden werden.**)). For the sake of clarity, the temporal summation over *m* is abbreviated by brackets [ ].

$$\begin{pmatrix} [x_{S,1S}\,x_{S,1S}]+[\Delta x_{S,1S}\,\Delta x_{S,1S}] & [x_{S,1S}\,x_{S,2S}] & \cdots & [x_{S1S}\,x_{S,nS}] & [x_{S,1S}\,x_{M,0M}] \\ [x_{S,2S}\,x_{S,1S}] & [x_{S,2S}\,x_{S,2S}]+[\Delta x_{S,2S}\,\Delta x_{S,2S}] & \cdots & [x_{S,2S}\,x_{S,nS}] & [x_{S,2S}\,x_{M,0M}] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ [x_{S,nS}\,x_{S,1S}] & [x_{S,nS}\,x_{S,2S}] & \cdots & [x_{S,nS}\,x_{S,nS}]+[\Delta x_{S,nS}\,\Delta x_{S,nS}] & [x_{S,nS}\,x_{M,0M}] \\ [x_{M,0M}\,x_{S,1S}] & [x_{M,0M}\,x_{S,2S}] & \cdots & [x_{M,0M}\,x_{S,nS}] & [x_{M,0M}\,x_{M,0M}]+[\Delta x_{M,0M}\,\Delta x_{M,0M}] \end{pmatrix}\cdot\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\\ \nu\end{pmatrix}=\begin{pmatrix}[x_{0S}\,x_{S,1S}]\\ [x_{0S}\,x_{S,2S}]\\ \vdots\\ [x_{0S}\,x_{S,nS}]\\ [x_{0S}\,x_{M,0M}]\end{pmatrix} \quad (3.18)$$

I.e. the second element in the first row of the matrix, $[x_{S,1S}\ x_{S,2S}]$, is equivalent to $\sum_{t_S=1}^{m}(\ x_S(p_1,t_S)\cdot x_S(p_2,t_S)\,)$. The data is transformed into anomalies w.r.t. the monthly mean ($\overline{x}$) and normalized by the standard deviation (σ_x) of three-hourly means of one month for each grid box (see WACMOS DDF, 2010 for more details).

Following Lindau and Schröder (2010), the errors are assumed to be random here, so that all mixed terms $[x_i \Delta x_i]$ are zero (no tendency of overestimating high and underestimating low values). Furthermore, all $[\Delta x_i \Delta x_j]$ are zero, if different observations are considered (no tendency to overestimate when the neighbour is overestimating). Here, the expression x_i is used substitutional for x, x_S and x_M. For satellite observations in particular the second assumption might be invalid because retrieval schemes tend to over- or underestimate under certain atmospheric conditions and then all pixels concerned would have the same tendency. However, here the assumption of random errors is kept, so that error variances occur only on the diagonal of the matrix (Lindau and Schröder, 2010).The task is to determine the weights **Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.**. Obviously, a solution is not possible for one single case. But if a time series of M measurements at each location pi is available, a reasonable requirement is that the mean squared deviation between prediction (**Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.**) and truth (x0) is minimal (Lindau and Schröder, 2010).

The merged SEVIRI+IASI water vapour end product provides tropospheric water vapour for three vertical layers (sfc - 850 hPa, 850 hPa - 500 hPa,and 500 hPa – 200 hPa) on almost full MSG disc over land. The domain is shown in Figure 2. The spatial sampling is $(0.25^{\circ})^2$. The instantaneous water vapour products from SEVIRI and IASI are retrieved under clear-sky conditions only.

4 Assumptions and limitations

Towards the application of climate monitoring of WV it is often recommended to use climatological background information. Both 1D-Var in their current setting are tuned towards the use of NWP or re-analysis fields. When running the schemes in ‘climate mode’, meaning that the background information is taken from climatological values, it is recommended to review the background error covariance matrix, since the associated error of climatological profiles as background are significantly higher than NWP or reanalysis fields. When not accounting for this fact, difficulties in converging as well as retrieval quality are likely to occur. Further, a implicit channel selection, as it is done in the IASI schemes under consideration of differences between simulated and measured radiances, will strongly depend on the background and/or first guess data.

For the IASI and SEVIRI instrument only a short common time span is available. Thus, the activities in GlobVapour does not create a climate time series but are very beneficial to investigate the quality of existing retrieval methods and to explore ways to combine the estimates into a higher value product. This is very ambitious as both instruments observe water vapour differently and have a different error characteristic.

The level of maturity of the sensor merged product will be assessed as a part of this project. However, there exist some reasons, why the maturity will probably be limited:

In the past, not much experience has been gained in the geostatistical combination of different sensor data that could guide the development of the new sensor merged products. Besides the difference in spatiotemporal samplings, differences in water vapour estimates from different spectral regions are present and not fully understood.

The bias correction within the merging scheme assumes higher quality for IASI than for SEVIRI. A confirmation for this assumptions is currently not available.

Preliminary validation results (WACMOS, 2010) exhibited promising quality of the recently released NWC-SAF software for the retrieval of water vapour from SEVIRI. As it can be expected from SEVIRI channel characteristics, lowest quality was found in the mid-troposphere. Further validation, in particular, over semi-arid to arid surfaces as present in the AMMA region (Niamey, Niger) need to be performed to finally decide, if the full disc SEVIRI+IASI product is feasible only for the upper troposphere (p<500 hPa).

Another limitation is that the level 2 water vapour products will only be generated under clear sky conditions. The filling of cloud gaps is an unresolved issue. Studies on the so called clear sky bias exist and indicate that cloudy areas should not be filled with values from neighboring clear sky pixels. Another issue, not dealt within this document, is the homogenization of the radiance time series. IASI and SEVIRI stability is not considered as critical due to the relatively short length of the time series. However, attention will be paid on this. Finally, more sophisticated ways of combining satellite observations at retrieval level could also be carried out but are beyond the scope of this study.

5 Conclusions

Two 1D-Var systems are implemented to retrieve the integrated water vapour vertically resolved in three layers bordered by the following pressure levels: 200hPa, 500hPa, 850hPa and surface. The derived WV is an optimal estimate considering the provided background information and the satellite measurements with their associated errors. To create the final IASI+SEVIRI product an objective analysis method is applied which delivers three-hourly and monthly means of the merged LCWV products defined over land and ocean surface within the SEVIRI disc.

The theoretical description as well as the practical application of the two schemes are described in this document. For further improvement of the schemes radiances bias corrections and improved representation of the background errors can be analysed. Such issues should be addressed in broader context and the project plans to support the GEWEX Radiation Panel in its efforts to start addressing these challenging issues in an international context.