Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 2

End to End ECV Uncertainty Budget (E3UB) - Part 2: CDR-3 & CDR-4

Ref: D2.3

Date: 27 June 2023

Issue: 4.0

For: ESA / ECSAT

Ref: CCIWV.REP.011

This Page is Intentionally Blank

Project	**:**	**Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 2**
Document Title:		**End to End ECV Uncertainty Budget (E3UB) - Part 2: CDR-3 & CDR-4**
Reference	**:**	**D2.3**
Issued	**:**	**27 June 2023**
Issue :		**4.0**
Client:		**ESA / ECSAT**

Author(s) : Michaela I. Hegglin (UoR), Hao Ye (UoR), Alexandra Laeng (KIT), Gabriele Stiller (KIT), Richard Siddans (RAL), Brian Kerridge (RAL), Kaley Walker (UoT), Chris Sioris (ECCC)

Document Change Log

Issue/ Revision	Date	Comment
1.0	12 June 2019	Initial issue
2.0	25 August 2020	Update with draft on CDR-3 and CDR-4 approach
2.1	13 October 2020	Update of version 2.0 according to RIDs
3.0	14 June 2022	Final issue (Phase 1)
4.0	27 June 2023	Scheduled update after year 1 of Phase 2

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

1. INTRODUCTION

1.1 Purpose

The purpose of this document is to deliver an end-to-end uncertainty budget that is based on estimates of uncertainties that arise from each step within the merging process chain of the climate data record, here for the deliverables CDR-3 and CDR-4. This document incorporates all changes made to the merging algorithm carried out during phase 1 and phase 2 of the WV_cci project.

1.2 Scope

The end-to-end uncertainty budget aims at including the uncertainties from the two processing steps (1) L1 to L2 (retrieval step) and (2) L2 to L3 (merging step). The retrieval step consists of uncertainties based on each step in the retrieval process including instrument noise characteristics, atmospheric corrections, geolocation and geophysical product retrieval. The merging step consists of uncertainties based on each step in the merging process chain, including the uncertainty arising from geophysical sampling biases when moving from L2 to L3 for each satellite product and the uncertainty in the use of external ancillary data (i.e. model quality used as transfer function between each satellite product).

2. VERTICALLY RESOLVED STRATOSPHERIC ZONAL MEAN CLIMATE DATA RECORD (CDR-3)

2.1 Stratospheric limb sounder L2 data uncertainties (CDR-3)

2.1.1 Characterisation of uncertainty estimates of WAVAS datasets

The L2 WAVAS-II satellite limb sounder datasets currently do not provide their bottom-up uncertainty estimates in a consistent manner, however each instrument team still provides uncertainty information for their datasets for each profile. A more detailed discussion of this information including examples is provided in [RD-1]. Table 2-1 summarises the ways the different satellite instruments from WAVAS-II activity report their uncertainties. It provides the information about major contributors to the uncertainties of Level 1 data out of which the dataset was produced. According to the ways the reported uncertainties of Level 2 are calculated, all WAVAS II satellite datasets are subdivided into following classes:

A: Uncertainties of the Level 2 data are dominated by the instrument noise propagated through the retrieval.

B: Uncertainties of the Level 2 data contain the noise propagated through the retrieval and other error components.

C: We assess the uncertainties of Level 2 data in a different way (not via error propagation of Level 1 data through the retrieval according to Rodgers).

Table 2-1: Major contributors to Level 1 uncertainties by instrument

Instrument	Major contributors to uncertainties of Level 1	A	B	C
ACE-FTS	Instrument line shape characterisation[1]			X[2]

[1] Currently this is not being characterised; we assume that these would be the major contributors.

[2] In the version provided, the statistical fitting error is reported as the uncertainty of Level 2.

Instrument	Major contributors to uncertainties of Level 1	A	B	C
ACE-MAESTRO	Tangent height registration, dark current, signal degradation over time[2]		X	
HALOE	Altitude registration, error due to aerosol			X
MIPAS-IMK	Spectral noise, gain (calibration error), ILS	X		
MLS	Radiance noise, calibration for long-term averages		X	
POAM III	Instrument absolute pointing information, [3] measurement noise, random altitude registration errors, radiance normalisation errors		X	
SAGE II	The L1 (i.e. slant-path transmission) are a combination of the uncertainties in both the unattenuated solar data (I-zero) and the attenuated atmospheric measurements (I) (T = I/I-zero). "I-zero" uncertainties are simply the standard error of those measurements. "I" uncertainties are based on the statistics of the data within each vertical bin (i.e. 0.5 km gridding)		X	
SAGE III M3M	Uncertainties in the unattenuated solar data, uncertainties in attenuated atmospheric measurements		X[4]	
SCIAMACHY limb	Instrument noise			X[5]

[3] Because water vapour is retrieved from the long wavelength channels, that are optically thin and hence very sensitive to small pointing errors.

[4] It is a combination of the propagated errors of the Level 1 optical depth, and estimated error associated with the attenuator etalon.

*[5] The dominating error is the instrument noise, it is propagated through the retrieval according to Rodgers, but we do not use the noise error from the level 1 product but instead the noise estimated from the residual scaled with 2/3. (SNR * 1.5)*

Instrument	Major contributors to uncertainties of Level 1	A	B	C
SMR	Side band response, instrument noise, non-linearity	X		

2.1.2 Uncertainty of instrument precision

No validation of the provided instrument precisions are currently available for the WAVAS-II datasets. This task is ongoing within the SPARC TUNER activity and will not be duplicated in the WV_cci project.

2.1.3 Uncertainties in seasonal cycle

The description of the data including uncertainties in the seasonal cycle can be found in [RD-2], with the values accessible at:
https://zenodo.org/record/2532028#.XOUb3_7grIU

2.2 L3 Data uncertainties SPARC Data Initiative climatologies

Since the overall uncertainty of the climatologies of each instrument are not known in a consistent way from bottom-up estimates for all the instruments (see [RD-1]), their L3 uncertainty is approximated by the standard error of the mean (***SEM*** or $\sigma_{\hat{x}}$) of an instrument (*i*) (Hegglin et al., 2013 [RD-3]):

$$\boldsymbol{SEM_i} = \boldsymbol{\sigma_{\hat{x}\_i}} = \sqrt{\frac{1}{N}}\,\boldsymbol{\sigma_{std\_i}} \qquad \textbf{(Equation 1)}$$

The $\boldsymbol{SEM_i}$ (or $\sigma_{\hat{x}\_i}$) for each instrument (*i*) can thereby readily be estimated from the standard deviation (σ_{std_i}) and number of measurements per bin (*N*), since these quantities are available in the SPARC Data Initiative WV climatology files submitted to Pangea (doi TBD). The standard deviation therein is defined over all measurements of a given instrument as:

$$\boldsymbol{\sigma_{std\_i}} = \sqrt{\frac{\sum_{n=1}^{N}(x_n - \langle x \rangle)^2}{N-1}} \qquad \textbf{(Equation 2)}$$

The x_n here denotes the n-th observation of a given instrument in a certain grid box with *N* being the total number of observations available in this grid box.

The $\boldsymbol{SEM_i}$ (or $\sigma_{\hat{x}\_i}$) reflects the influence of random errors but does not include systematic errors (bias). The bias of the instrument mean will be considered in the next step (see Section 2.3).

Uncertainties in the SPARC Data Initiative WV climatologies that arise from sampling biases (i.e. due to a non-uniform temporal sampling of a given month or an uneven latitudinal and/or longitudinal spatial sampling [RD-4]) have been estimated based on sub-sampling of a CCM by Toohey et al. (2013) [RD-4]. Differences in the averaging technique used to produce the zonal mean climatologies [RD-5] will also be considered by using the rule of error propagation to extend the SEM_i to σ_{i_ext} according to:

$$\boldsymbol{\sigma_{i\_ext}} = \sqrt{(\boldsymbol{\sigma_{\hat{x}\_i}})^2 + (\boldsymbol{\sigma_{samp\_i}})^2 + (\boldsymbol{\sigma_{ave\_i}})^2} \qquad \textbf{(Equation 3)}$$

2.3 L3 Data uncertainties CDR-3

We are utilising the L3 SPARC Data Initiative uncertainty estimates given in Section 2.2 as main input to the uncertainty calculation of CDR-3. The variables listed in the following are added to the CDR-3 VRWV file content in addition to the overall uncertainty.

- **Standard deviation** (over all bias-corrected instrument means going into a grid point):

$$\boldsymbol{\sigma_{std}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \langle x \rangle)^2}{N-1}} \qquad \textbf{(Equation 4)}$$

  The x_i here denotes the mean value of the i-th instrument in a certain grid point with N being the total number of instruments and $\langle x \rangle$ the multi-instrument mean over these mean values. Note, this variable is not available for grid points where only one instrument is providing information. In such cases, the user is advised to use the propagated uncertainty of the instrument's data point (see Section 2.2) to get a likely range of possible values.

- **The uncertainty measure** at each grid point in space and time is derived from the optimal estimation (OE) theory and was used to minimise the originally planned propagated uncertainty.

  The OE thereby helps optimising the weights that need to be applied to each instrument, with the resulting estimated uncertainty expressed by

$$\boldsymbol{\sigma_{OE}} = \left(\sqrt{\sum_{i=1}^{N} \frac{1}{\sigma_{\hat{x}\_i}^2}} \right)^{-1} \qquad \textbf{(Equation 5)}$$

This approach will need to go through a peer-review process, with a journal publication on the methodology in preparation.

3. VERTICALLY RESOLVED UPPER TROPOSPHERIC/LOWER STRATOSPHERIC CLIMATE DATA RECORD (CDR-4)

3.1 Stratospheric limb sounder L2 data uncertainties (CDR-4)

The L2 limb sounder uncertainties for CDR-4 are same as the Section 2.1.1 WAVAS dataset for CDR-3. In the production of CDR-4, limb sounder datasets MLS and MIPAS are used, in which the MIPAS data is obtained from the WAVAS-II dataset. The new v5.0 MLS was downloaded directly from the official data website and includes the L2 data uncertainties.

3.2 IMS L2 data uncertainties (CDR-4)

3.2.1 Estimated uncertainty based on diagnostics reported at L2

The RAL Infra-red Microwave Sounder (IMS) scheme is described in the ATBD [RD-7]. This is an optimal estimation (OE) algorithm which provides a number of diagnostics useful for characterising the uncertainty in the L2 product. It is assumed that the reader is familiar with the ATBD and in particular section 2.8 which describes how individual profile uncertainties are estimated and reported at L2, together with averaging kernels which characterise the retrieval sensitivity. Here we mainly present an overview of the uncertainties reported at L2 and assess their realism by comparing profiles to ECMWF (ERA interim) analysis. An assessment of the water vapour profile sensitivity to cloud is also reported.

It should be noted that the IMS scheme simultaneously retrieves water vapour together with many other parameters, including surface and atmospheric temperature profile, surface spectral emissivity, ozone profile and scale factors for two systematic spectral residual patterns. Furthermore measurement errors are estimated from comparison of forward model (FM) simulations based on realistic predictions of the atmospheric state (from the EUMETSAT version 6 piece-wise linear regression scheme). The assumed measurement errors therefore represent random forward model errors as well as instrument noise. This means that many potential sources of random error are included in the estimated solution covariance which is reported at L2, including e.g. instrument noise, errors due to uncertainty in surface emissivity, temperature profile, FM errors etc. Errors that remain difficult to quantify are systematic errors which might, for example, arise from errors in spectroscopy or errors which the FM cannot represent

(e.g. unknown spectral features in surface emissivity). Bounds on these errors can be estimated by comparing retrievals to independent data (e.g. ERA-interim analysis as shown here or the sonde and ground-based data used in the CCI+ validation work).

Some basic retrieval diagnostics from the IMS version 1 data are presented for three typical scenes in Figure 3-1 to Figure 3-3. In each case, the left-hand panel shows a comparison of the retrieved (Ret, black), *a priori* (AP, red), ECMWF (ERA interim) analysis (NWP, green) and ECMWF analysis after applying averaging kernels as described in the ATBD (NWPxAK, blue). Dashed lines about the retrieved profile indicate the range of the estimated standard deviation (ESD) from the solution covariance. Right-hand panels show averaging kernels for a regularly sampled subset of the 101 output retrieval levels (each line showing the sensitivity of the given output level to perturbations in the true profile, as a function of altitude).

These examples illustrate the capability of IMS to resolve water vapour throughout the troposphere and return profiles close to NWP, even where this deviates strongly from the climatological prior. The altitude up to which the water vapour profile is well resolved follows the tropopause: sensitivity to stratospheric water vapour is generally low (due mainly to the relatively low stratospheric mixing ratio).

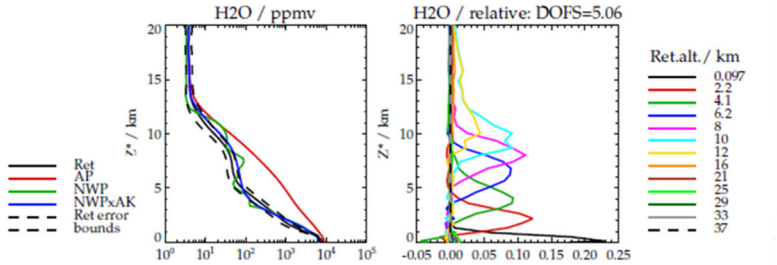

Figure 3-1: Retrieval diagnostics for water vapour for a mid-latitude scene over sea.

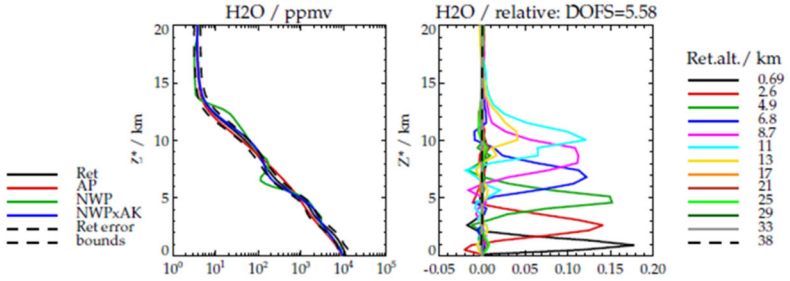

Figure 3-2: Retrieval diagnostics for water vapour for a mid-latitude scene over land.

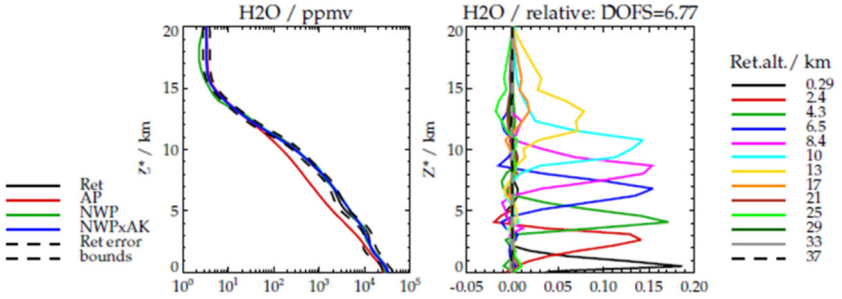

Figure 3-3: Retrieval diagnostics for water vapour for a tropical scene over land.

The global variation of retrieval quality is summarised in Figure 3-4. This shows results averaged over 10-degree latitude bins for a particular month (July 2008). The variation of some of these diagnostics over the mission is illustrated in section 4.2 of the ATBD. Here we seek to illustrate the typical variation of individual retrievals for which a monthly zonal analysis is sufficient. Retrieval quality is somewhat different over land and sea, mainly due to the strong land/sea contrast in microwave emissivity which generally leads to better near surface sensitivity over sea. The level and nature of cloud cover in scenes for which a v1 IMS retrieval is produced is also somewhat different for land vs sea, (due to the nature of the pre-retrieval cloud screening based on a simple

brightness temperature difference technique, see also Section 3.2.2). Results are therefore presented in pairs of panels for scenes over sea and land, respectively. The figure shows the following information:

- Left-hand pairs of panels (from top to bottom):
  - The assumed *a priori* profiles
  - The mean retrieved profiles
  - The mean co-located “NWP” profile (ECMWF ERA-interim analysis)
  - NWP after averaging kernels are applied to it to simulate the retrieval smoothing (“NWPxAK”)
  - The standard deviation (SD) of the individual retrieved profiles within each latitude bin
  - The SD of the NWP profiles
  - The SD of the NWPxAK profiles
  - The diagonal values of the square averaging kernel (AK) matrix, normalised by the associated layer thickness. These can be integrated with respect to altitude to give the total degrees of freedom. They also correspond to the inverse of the vertical resolution in km (i.e. where the value is 1 the vertical resolution approaches 1 km; more typical values of 0.5 indicate a vertical resolution of around 2 km).
- Right-hand pairs of panels show the following (from top to bottom):
  - The mean relative difference between retrieval and the NWP profiles (in per cent)
  - The mean relative difference between retrieval and NWPxAK
  - The standard deviation of the individual profile differences between retrieval and NWP
  - The standard deviation of the individual profile differences between retrieval and NWPxAK
  - The assumed *a priori* error profile (this is the same for all scenes)
  - The estimated standard deviation (ESD) of the total retrieval error
  - The estimated retrieval noise. Note the difference between the total variance (i.e. the square of the ESD) and the noise variance would give the retrieval smoothing error.

The following key points can be made:

- The retrieval is generally sensitive up to the tropopause. Above that the AK diagonal values become very low, the ESD tends to the prior error and noise tends to zero (all indicating the retrieval tends to the prior in the stratosphere). This is reflected in the better agreement between retrieval and NWPxAK than with NWP, in particular above the tropopause.

- Within the troposphere the estimated total (random) uncertainty (from the ESD) is estimated to be around 20–40%, with a contribution from noise of 10–15%.
- The estimated errors can be compared with the standard deviation in the difference between NWP and retrieval. Ideally, if the model perfectly represents the truth and the *a priori* error covariance is consistent with the true variability, then the standard deviation in differences between retrieval and NWP should be similar to the ESD and the standard deviation of differences between retrieval and NWPxAK should be similar to the estimated noise. In practice model and (time+space) sampling errors will mean ESD and noise should underestimate the actual standard deviations. This is indeed seen to be the case: in particular, there is a reasonably good correspondence between the ESD and the standard deviation of differences between NWP and retrieval, indicating that the ESDs provide a generally reasonable estimate of the total uncertainty. However differences can be larger over very cold land surfaces (up to 70%) or over desert (the peak round 30 N in the land SD).
- Systematic (mean) differences between retrieval and NWPxAK give an indication of systematic errors in the IMS retrieval (assuming the model to be correct). Relative differences are largest in the tropical and mid-latitude mid-upper troposphere, where IMS is biased low compared to NWPxAK by up to ~25%.

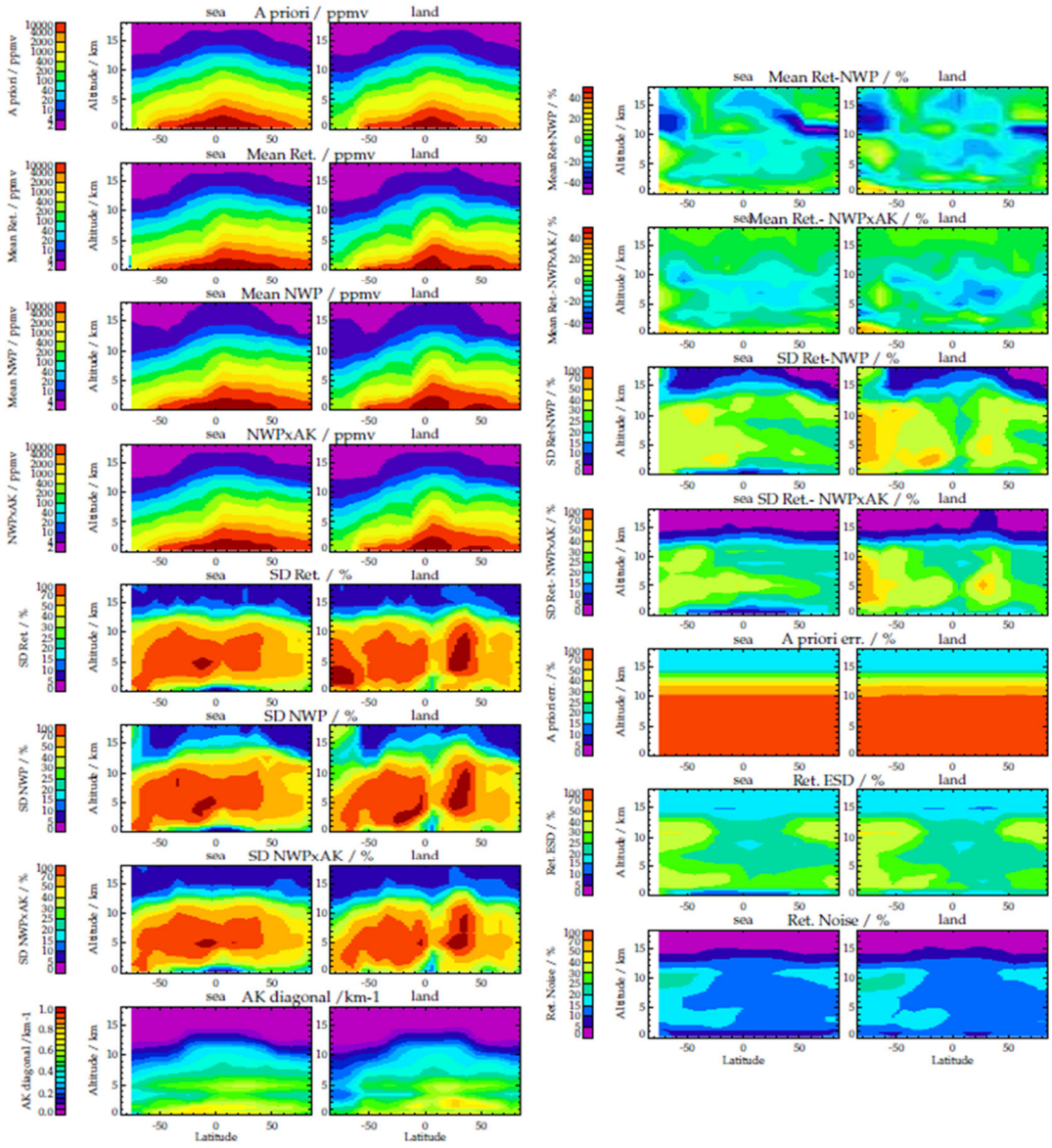

Figure 3-4: Illustration of zonal mean water vapour retrieval diagnostics from IMS (considering all data in July 2008). Results are shown as latitude vs height zonal cross-sections, in pair of panels for scenes over sea (left of each pair) and land (right of each pair). Zonal means are constructed in 10-degree latitude bins. See text for an explanation of the various panels.

3.2.2 Effect of cloud

In the v1 IMS scheme, a brightness-temperature (BT) difference test is used to pre-screen IASI scenes before running the retrieval. This test removes scenes with a high fraction of high altitude and/or optically thick cloud. There remain many scenes with partial, low-altitude or optically thin cloud cover. IMS retrieves an effective cloud fraction and height to accommodate (some of) the impact of this cloud, however errors on the

water vapour profile will be increased, and particularly at altitudes below the cloud top. Here we illustrate the impact of cloud on the retrieved profiles by analysing differences between retrieval and ERA-interim as a function of the co-retrieved effective cloud fraction and height. These parameters can be used to filter L2 data for adversely affected scenes.

Results are illustrated in Figure 3-5. As before, results are shown separately in pairs of panels for sea and land scenes, considering all retrievals in July 2008):

- Top left-hand panels: Two dimensional histograms of the base-10 logarithm of the number of scenes in each (regularly spaced) bin of cloud-fraction and height. The area of the plotted rectangles indicates the extent of each bin. White regions indicate no retrievals in the bin. Note that the BT difference test effectively removes high altitude, high fraction cloud, so scenes that remain are either low altitude or low fraction. Somewhat higher fractions of relatively high altitude cloud are apparent in land scenes. All other panels show the mean of another quantity considering all the scenes in each cloud fraction/height bin.
- 2nd from top: Mean retrieval cost function value (meaning the quality of the fit to observed radiances) in each bin. Note that scenes towards the high altitude/fraction edge of the histogram tend to have slightly higher cost, indicating that the cloud degrades the (average) fit quality. Some of the high cost function values over land are unrelated to cloud: scenes over desert also give rise to relatively high cost due to current limitations in the modelling of desert surface spectral emissivity.
- Other panels show relative differences between the retrieval and ERA interim at a specific vertical level (indicated in the plot title), averaged over the cloud fraction/height bins. Pairs of panels on the left show differences accounting for retrieval smoothing using the averaging kernels (NWPxAK), while pairs on the right show the direct comparison to ERA-interim (NWP).

The averaging kernels are computed including the effect of the fitted cloud in the radiative transfer. It is seen that, as expected, using the averaging kernels improves the agreement with ERA-interim at most levels, under most cloud conditions. However, there remain significant errors at levels below the fitted cloud, when the cloud fraction is relatively high (e.g. see red area in 4.1-km panel over land). Restricting results to where either cloud fractions < 0.1 or cloud height < 2 km would effectively remove scenes apparently biased by cloud. Alternatively, scenes can be effectively filtered using the product of cloud fraction * height < 1 km.

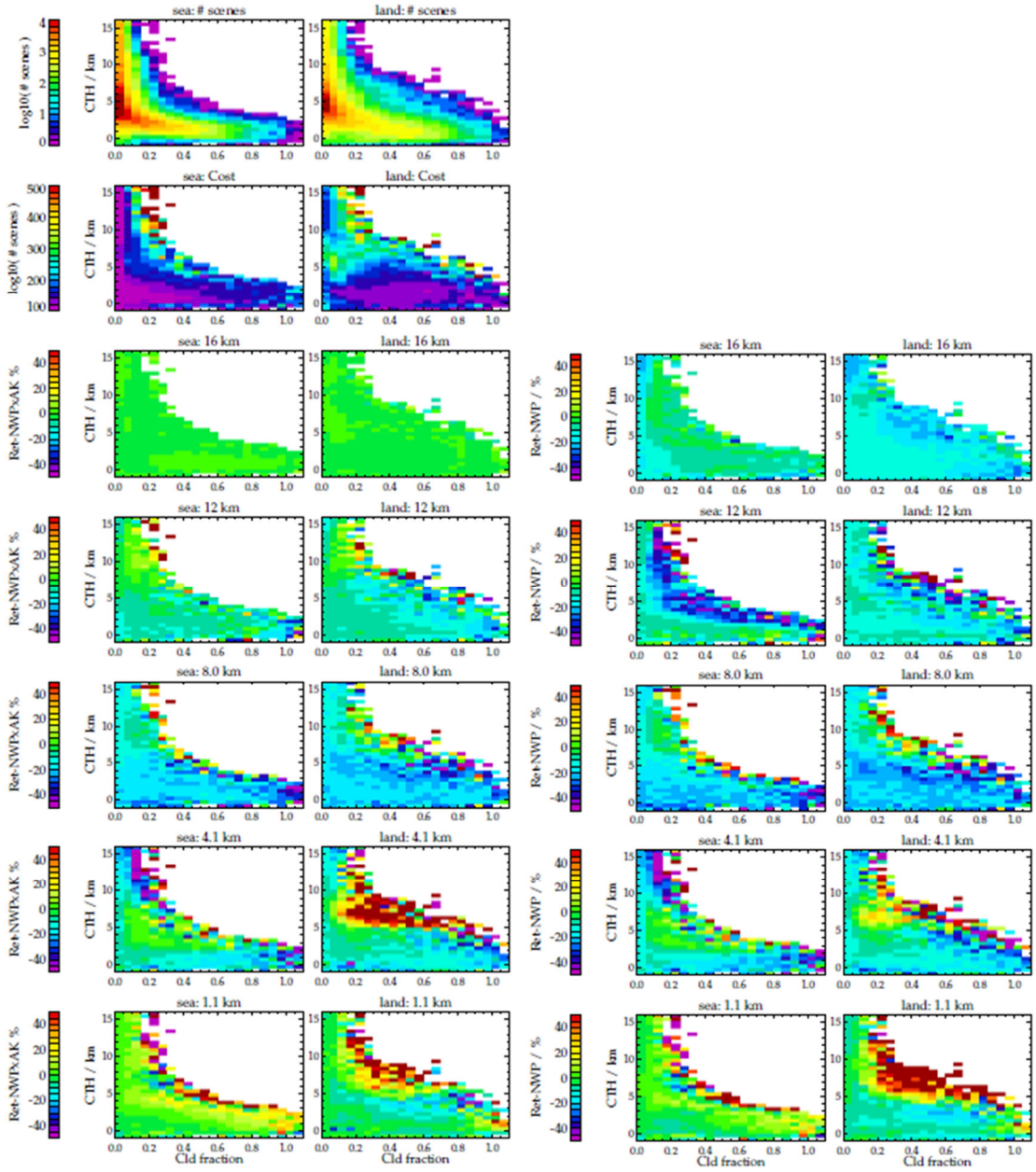

Figure 3-5: Mean relative differences between retrieved water vapour (at various levels) and ERA interim as a function of (co-retrieved) cloud fraction and height. Results are shown for all profiles in July 2008, separately for land and sea in pairs of panels. Panels in the top left show the number of scenes in each cloud-fraction/cloud height bin. Panels below that show the mean relative difference between retrieval and NWP at a specific vertical level (indicated in the panel heading); panel pairs on the left show differences with respect to NWPxAK; panel pairs on the right show corresponding differences with respect to NWP (without AK).

3.2.3 Cross-track dependence of retrievals

Figure 3-6 shows an analysis of the cross-track dependence of the water vapour retrieval as a function of across-track scan position. Panels are similar to those in Figure 3-5, except that in this case the relative differences between the retrieval and ERA interim are averaged as a function of across-track scan index and 10-degree latitude bins. Note that IASI measures four fields-of-view at 30 across-track scan positions, giving 120 across-track samples in total. The figure shows that biases present in direct comparison to ERA Interim (NWP) are largely explained by using the averaging kernels (NWPxAK) and that these biases have a small dependence on cross-track position. There is some indication of a tendency towards negative bias towards the edge of the swath, particularly at low altitude.

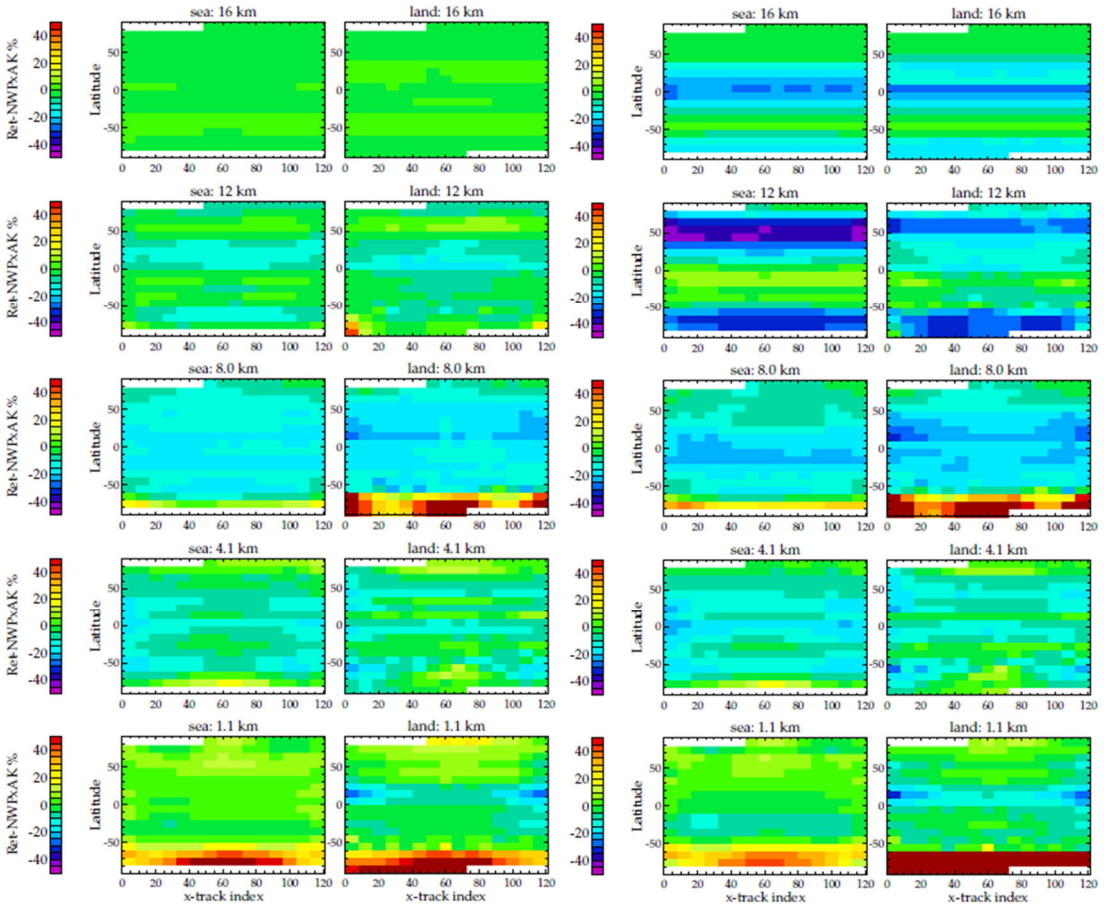

Figure 3-6: Mean differences between retrieved water vapour and ERA interim averaged as a function of across-track scan index and latitude, for specific altitude (from top to bottom). Results are shown for all profiles in July 2008. Pairs of panels on the left compare the retrieval to NWPxAK, panels on the right to NWP.

3.3 L3 Data uncertainties (CDR-4)

3.3.1 L3 data uncertainties for each instrument

The monthly level 3 data for all the instruments are calculated from the individual level 2 data for each L3 bin and the estimated uncertainty provided is calculated from the spread of L2 data in each L3 bin. The following variables are available for each instrument:

- **Mean value** (for each instrument within a grid point):

$$\langle x_i \rangle = \frac{\sum_{n=1}^{N} x_n}{N} \qquad \textbf{(Equation 6)}$$

The x_n here denotes the n-th measurement of the i-th instrument in a certain grid point with *N* being the total number of measurements. It should be noted that the IMS data is provided with value as logarithm of the mixing ratio and the mean value is also averaged in the logarithm of the mixing ratio.

- **Standard deviation**:

$$\boldsymbol{\sigma_{std\_i}} = \sqrt{\frac{\sum_{n=1}^{N}(x_n - \langle x_i \rangle)^2}{N-1}} \qquad \textbf{(Equation 7)}$$

Here, the $\langle x_i \rangle$ is the mean value of the *N* measurements of the i-th instrument.

- **Standard error of the mean (SEM):**

$$\boldsymbol{SEM_i} = \boldsymbol{\sigma_{\hat{x}\_i}} = \sqrt{\frac{1}{N}}\,\boldsymbol{\sigma_{std\_i}} \qquad \textbf{(Equation 8)}$$

The $\boldsymbol{SEM_i}$ here denotes SEM for the i-th instrument in a L3 bin with *N* being the total number of measurements.

- **Mean uncertainty** $\langle \boldsymbol{\sigma_i} \rangle$:

$$\langle \boldsymbol{\sigma_i} \rangle = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\sigma_{n\_i}} \qquad \textbf{(Equation 9)}$$

The $\boldsymbol{\sigma_{n\_i}}$ here denotes the uncertainty of the n-th measurement of the i-th instrument in a certain grid point.

- **Mean of the squares of the uncertainty** $\langle \boldsymbol{\sigma_i^2} \rangle$:

$$\langle \boldsymbol{\sigma_i^2} \rangle = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\sigma_{n\_i}^2} \qquad \textbf{(Equation 10)}$$

The $\langle \boldsymbol{\sigma}_i^2 \rangle$ here is provided as the mean uncertainty for the i-th instrument at a certain grid point. Note that the mean uncertainty for IMS here is calculated from the L2 uncertainties of the logarithm of mixing ratio.

For the L3 data uncertainties of each instrument, one critical term that should be noted are the uncertainties from the quantile mapping bias correction. First, the normal distribution fitting applied to the reference dataset (balloon-borne hygrometer data here) introduces uncertainty into the bias correction process, which strongly depends on the amount of data point used in the fitting. Second, the uncertainty of the quantiles from the fitted distribution of reference data is related to the quantile values and the sampling size [RD-8]. The bias correction from the quantile–quantile matching between satellite observations and reference data also brings in new uncertainties propagated from the reference data. The uncertainty of the quantiles is expected to be much smaller than the uncertainties from the satellite observations [RD-8] and the uncertainty from bias correction is neglected for CDR-4.

3.3.2 L3 data uncertainties for merged product CDR-4

The merged CDR-4 product is created from several satellite instruments and the estimated uncertainties from Section 3.3.1 are utilised to estimate the uncertainty information for the CDR-4 product.

- **Mean value** (over all instruments going into a grid point):

$$\langle \mathrm{x} \rangle = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} \mathrm{x}_{\mathrm{i}}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}}} \qquad \textbf{(Equation 11)}$$

  The x_i here denotes the mean value of the i-th instrument and the α_i here denotes the weight of the i-th instrument in the merged mean in a certain grid point with *N* being the total number of instruments.

- **Standard deviation** (over all instrument means going into a grid point):

$$\boldsymbol{\sigma}_{\mathbf{std}} = \sqrt{\frac{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} (\mathbf{x}_{\mathbf{i}} - \langle \mathbf{x} \rangle)^2}{\mathbf{N}-\mathbf{1}}} \qquad \textbf{(Equation 12)}$$

- **Mean uncertainty** $\langle \boldsymbol{\sigma} \rangle$ (based on the $\mathrm{SEM_i}$ or $\sigma_{\hat{x}\_i}$ from the L3 data for each instrument):

$$\langle \boldsymbol{\sigma} \rangle = \frac{\mathbf{1}}{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \boldsymbol{\alpha}_i} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \boldsymbol{\alpha}_i \boldsymbol{\sigma}_{\hat{x}\_i} \qquad \textbf{(Equation 13)}$$

- **Propagated uncertainty** (based on the SEM or $\sigma_{\hat{x}}$ from the L3 data for each instrument):

$$\boldsymbol{\sigma_{prop}} = \frac{\mathbf{1}}{\sum_{\mathbf{i=1}}^{\mathbf{N}} \boldsymbol{\alpha_i}} \sqrt{\sum_{\mathbf{i=1}}^{\mathbf{N}} \boldsymbol{\alpha_i^2 \sigma_{\hat{x}\_i}^2}} \qquad \textbf{(Equation 14)}$$

This propagated uncertainty in principle can be minimised using the optimal estimation (OE) theory. The OE helps optimising the weights that need to be applied to each instrument as

$$\boldsymbol{\alpha_i} \propto \frac{\mathbf{1}}{\boldsymbol{\sigma_{\hat{x}\_i}^2}}, \qquad \textbf{(Equation 15)}$$

with the resulting estimated uncertainty expressed by

$$\boldsymbol{\sigma_{OE}} = \left(\sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \frac{\mathbf{1}}{\boldsymbol{\sigma_{\hat{x}\_i}^2}}} \right)^{\mathbf{-1}}. \qquad \textbf{(Equation 16)}$$

The most critical term in the uncertainty of the merged CDR-4 product is the choice of weights for the satellite instruments. Currently, the choice of weights is based on the OE theory and using the uncertainty from each instrument. Due to the spatial and temporal coverage, the uncertainty of L3 data from each instrument is sensitive to the total amount of available observations, considering the difference between the two kinds of satellite viewing geometry. Further studies on the representativeness and spatial variability for each satellite data will provide more appropriate weights for the calculation of merged CDR-4 product and reduce the uncertainty in the final product.

APPENDIX 1: REFERENCES

[RD-1] ESA Water Vapour cci: D2.1 Product Validation and Selection Report (PVASR), v2.2, 17 August 2021.

[RD-2] Lossow, S., Khosrawi, F., Nedoluha, G. E., Azam, F., Bramstedt, K., Burrows, John. P., Dinelli, B. M., Eriksson, P., Espy, P. J., García-Comas, M., Gille, J. C., Kiefer, M., Noël, S., Raspollini, P., Read, W. G., Rosenlof, K. H., Rozanov, A., Sioris, C. E., Stiller, G. P., Walker, K. A., and Weigel, K.: The SPARC water vapour assessment II: comparison of annual, semi-annual and quasi-biennial variations in stratospheric and lower mesospheric water vapour observed from satellites, Atmos. Meas. Tech., 10, 1111-1137, https://doi.org/10.5194/amt-10-1111-2017, 2017.

[RD-3] Hegglin, M. I., et al. (2013), SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, J. Geophys. Res. Atmos., 118, 11,824–11,846, doi:10.1002/jgrd.50752.

[RD-4] Toohey, M., et al. (2013), Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res. Atmos., 118, doi:10.1002/jgrd.50874.

[RD-5] Funke, B., and T. von Clarmann (2012), How to average logarithmic retrievals?, Atmos. Meas. Tech., 5, 831–841.

[RD-6] ESA Water Vapour cci: D2.2 Algorithm Theoretical Basis Document (ATBD) Part 3 – CDR-3 and CDR-4, Issue 2.0, 7 June 2022.

[RD-7] ESA Water Vapour cci: D2.2: Algorithm Theoretical Basis Document (ATBD) Part 2 - IMS L2 Product; Issue 1.0, 27 March 2019.

[RD-8] Matsuyama, T. (2019): Estimation of uncertainty of percentile values in particle size distribution analysis as a function of number of particles, Advanced Powder Technology, 30, 2616–2619, doi:10.1016/j.apt.2019.08.008.

APPENDIX 2: GLOSSARY

Term	Definition
ACE–FTS	Atmospheric Chemistry Experiment Fourier Transform Spectrometer
ACE–MAESTRO	Atmospheric Chemistry Experiment Measurements of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation
AK	Averaging Kernel
ATBD	Algorithm Theoretical Basis Document
BT	Brightness Temperature
CCI	Climate Change Initiative
CCM	Chemistry–Climate Model
CDR	Climate Data Record
E3UB	End to end Uncertainty Budget
ECMWF	European Centre for Medium-range Weather Forecasts
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
ERA-Interim	ECMWF Re–Analysis Interim
ESA	European Space Agency
ESD	Estimated Standard Deviation
FM	Forward Model
GOMOS	Global Ozone Monitoring by Occultation of Stars
HALOE	Halogen Occultation Experiment
HIRDLS	High Resolution Dynamics Limb Sounder
IASI	Infrared Atmospheric Sounder Interferometer
IFOV	Instantaneous Field of View
ILAS–II	Improved Limb Atmospheric Spectrometer–II
ILS	Instrument Line Shape
IMS	Infra-red Microwave Sounder
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
NICT	National Institute of Information and Communication Technology
NWP	Numerical Weather Prediction

Term	Definition
OE	Optimal Estimation
POAM	Polar Ozone and Aerosol Measurement
SAGE	Stratospheric Aerosol and Gas Experiment
SAGE III M3M	SAGE III Meteor-3M
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY
SD	Standard Deviation
SMILES	Solar wind Magnetosphere Ionosphere Link Explorer
SEM	Standard Error of the Mean
SMR	Submillimeter Wave Radiometer
SOFIE	Solar Occultation For Ice Experiment
SPARC	Stratosphere–troposphere Processes And their Role in Climate
VRWV	Vertically Resolved Water Vapour
WAVAS	WAter Vapour ASsessment
WV	Water Vapour

End of Document