ESA Cloud_cci

Comprehensive Error Characterisation Report

Issue 4 Revision 1

30 April 2018

Deliverable No.:	D-1.3
ESRIN/Contract No.:	4000109870/13/I-NB
Project Coordinator:	Dr. Rainer Hollmann Deutscher Wetterdienst rainer.hollmann@dwd.de
Technical Officer:	Dr. Simon Pinnock European Space Agency pascal.lecomte@esa.int

Document Change Record

Document, Version	Date	Changes	Originator
1.0	09/03/2012	Version for ESA approval	C. Poulsen, R. Hollmann, M.Stengel
1.1	05/09/2012	Minor modifications after ESA review	N. Schneider
2.0 to be submitted to ESA	15/10/2013	Updated with results of AATSR/ORAC uncertainty validation (Section 8). Including results of optional work packages WP 9083 in Section 5 as paragraph "Uncertainties in statistical models" Editorial adjustments	C. Poulsen, L. Gómez Chova, M. Stengel
Version 2.0	08/11/2013	Internal review, editorial updates	O. Sus R. Hollmann
Version 3, submitted	15/05/2015	Adopting front page to Phase 2. Editorial finalization.	M. Stengel
Version 3.1 approved version	17/07/2015	Minor modifications after ESA review. Major editing.	C. Poulsen, C. Schlundt
Version 4.0 submitted version	22/02/2017	- few modifications to Table 3-1 - extending Section 4.2 - Complete revision of Section 6 (Uncertainty in/of Level-3 products) - Significant modifications to Section 7 - Removing old Section 9 - Changes to the conclusions section	M. Stengel C. Poulsen
Version 4.1 approved version	16/03/2018	Revision according to ESA review, e.g.: - Adding subsection 7.1.2 on FAME-C uncertainty validation - Adding subsection 7.2 on multi-layer cloud conditions	M. Stengel C. Poulsen S. Testorp

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Purpose

The Cloud_cci project produces long time series of cloud properties from the AVHRR, ATSR, and MODIS series of passive, polar-orbiting instruments. The data sets aim to have high stability and have been optimised to apply the latest instrument calibrations. The retrieval algorithms use mathematically rigorous techniques to retrieve products and associated uncertainties. This document is intended to inform the user of the current limitations of and uncertainties associated with the suite of cloud products produced. It also aims to assist users in utilising the product's uncertainties for their own analyses.

1.2 Reference Documents

[R-1] Heidinger, A.K. and Stephens, G.L., 1998. Nadir sounding of clouds and aerosols in the O2 A-band. Atmospheric science paper; no. 650.

[R-2] Siddans, R., Poulsen, C. and Carboni, E., 2010. Cloud Model for operational retrievals from MSG-SEVIRI, Eumetsat Final report. Eur. Organ. for the Exploit. of Meteorol. Satell., Darmstadt, Germany.

[R-3] Kratz, D.P., Chou, M.D., Yan, M.M.H. and Ho, C.H., 1998. Minor trace gas radiative forcing calculations using the k distribution method with one-parameter scaling. Journal of Geophysical Research: Atmospheres, 103(D24), pp.31647-31656.

[R-4] Cooper, S.J., L'Ecuyer, T.S., Gabriel, P., Baran, A.J. and Stephens, G.L., 2006. Objective assessment of the information content of visible and infrared radiance measurements for cloud microphysical property retrievals over the global oceans. Part II: Ice clouds. Journal of applied meteorology and climatology, 45(1), pp.42-62.

[R-5] Sayer, A.M., Thomas, G.E. and Grainger, R.G., 2010. A sea surface reflectance model for (A) ATSR, and application to aerosol retrievals.

[R-6] Round Robin evaluation protocol of the ESA Cloud_cci project (www.esa-cloud-cci.org)

[R-7] PVASR – Product Validation and Algorithm Selection Report of the ESA Cloud_cci project (www.esa-cloud-cci.org)

[R-8] Maddux, BC, Ackerman, SA, Platnick, S (2010). Viewing Geometry Dependencies in MODIS Cloud Products. *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY*, 27(9), 1519-1528.

[R-9] Poulsen, C. A., Siddans, R., Thomas, G. E., Sayer, A. M., Grainger, R. G., Campmany, E., Dean, S. M., Arnold, C., and Watts, P. D.: Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR, Atmos. Meas. Tech., 5, 1889-1910, https://doi.org/10.5194/amt-5-1889-2012, 2012.

[R-10] Stengel, M., Mieruch, S., Jerg, M., Karlsson, K. G., Scheirer, R., Maddux, B., ... & Hollmann, R. (2015). The Clouds Climate Change Initiative: Assessment of state-of-the-art cloud property retrieval schemes applied to AVHRR heritage measurements. Remote Sensing of Environment, 162, 363-379.

[R-11] GEWEX Cloud Assesment report Stubenrauch, C., Rossow, W., Kinne, S., and the GEWEX cloud assessment group 2012: Assessment of Global Cloud Datasets from Satellites report http://climserv.ipsl.polytechnique.fr/gewexca/papers/GEWEX_CA_2012.pdf

[R-12] Error propagation in the generation of gridded satellite datasets ,DWD Visiting Scientist Report, Ralf Bennartz, University of Wisconsin - Madison

[R-13] Saunders et al. RTTOV Science and validation report http://research.metoffice.gov.uk/research/interproj/nwpsaf/rtm/docs_rttov10/users_guide_10_v1.3.pdf

[R-14] Kratz D.P The correlated k-distribution technique as applied to AVHRR channels. Journal of Quantitative Spectroscopic Radiation Transfer 53, 501-517, 1995.

Heidinger A.K. and G.L. Stephens Nadir sounding of clouds and aerosols in the 02 A band PhD dissertation 225p Dep. Of Atmos. Sci. Colo. State Univ. Fort Collins 1998.

[R-15] Liu, J., C. Schaaf, A. Strahler, Z. Jiao, Y. Shuai, Q. Zhang, M. Roman, J. A. Augustine, and E. G. Dutton, Validation of Moderate Resolution Imaging Spectroradiometer (MODIS) albedo retrieval algorithm: Dependence of albedo on solar zenith angle, J. Geophys. Res., 114, D01106, doi:10.1029/2008JD009969.,2009

[R-16] Guenther, B., and Coauthors, 1996: MODIS Calibration: A Brief Review of the Strategy for the At-Launch Calibration Approach. J. Atmos. Oceanic Technol., 13, 274–285. doi: 10.1175/1520-0426. 1996

[R-17] Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., Christensen, M., Carbajal Henken, C., Preusker, R., Fischer, J., Devasthale, A., Willén, U., Karlsson, K.-G., McGarragh, G. R., Proud, S., Povey, A. C., Grainger, R. G., Meirink, J. F., Feofilov, A., Bennartz, R., Bojanowski, J. S., and Hollmann, R.: Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project, Earth Syst. Sci. Data, 9, 881-904, https://doi.org/10.5194/essd-9-881-2017, 2017.

[R-18] McGarragh, G. R., Poulsen, C. A., Thomas, G. E., Povey, A. C., Sus, O., Stapelberg, S., Schlundt, C., Proud, S., Christensen, M. W., Stengel, M., Hollmann, R., and Grainger, R. G.: The Community Cloud retrieval for CLimate (CC4CL). Part II: The optimal estimation approach, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-333, in review, 2017.

1.3 Structure of the document

The document begins by defining its terminology and describing the sources of uncertainty (Section 2 and 3). Section 4 reviews the methodologies used to determine the uncertainties and Section 5 describes the validation of the uncertainties in the products. After that, an overview of the uncertainty characterization in the level 3 products is given (Section 6), followed by cloud parameter uncertainty validation (Section 7) and a summary of the documentation and guidelines provided with the products (Section 8). Section 9 concludes this document.

2. Definition of terms

The definition of Error, Uncertainty, Uncertainty information, Uncertainty characterization, Validation, Accuracy, Precision, Stability, Representativity, Error co-variance matrix follows those given in the ESA Climate Change Initiative, Guidelines on Uncertainty characterisation document (Appendix 1).

3. Sources of uncertainty

The sources of uncertainty associated with the CC4CL Cloud_cci product can be categorised into a small number of subgroups:

a. Uncertainty in the forward model;

b. Uncertainty in the instrument;

c. Uncertainty in auxiliary data sets.

Table 3-2 outlines the current sources of uncertainty identified in the Cloud_cci products. Wherever possible, the underlying type of error (systematic or random) and its impact (small vs. large) is indicated. References to papers or reports where the error has been evaluated are included. At the current time there is still some ambiguity related to some of these sources. There may be additional sources of uncertainty which we have not yet envisaged. The table will be updated with new information as it becomes available.

It should be noted that although there are a significant number of uncertainties in the cloud products, this is to our best knowledge the first time that such an analysis of uncertainties in cloud products has been examined in detail. The uncertainties listed here are common to most cloud retrievals of similar genre. While they do not make the analysis and use of products any easier, the confidence in analysis resulting from these projects should have a much higher confidence if they have been understood.

FM Forward model error
APE Apriori error
SIM Assess error via simulations or other means and report separately
RAN Random error
SYS Systematic error
Priority
1 : Most important - essential to quantify
2 : Should be quantified with project
3 : Minor impact but should be qualified for completeness
4: Least important - only quantify if time

Table 3-1 *List of acronyms used in Table 3-2.*

Significance	Uncertainty	Size	Reference	Mechanisms	type	How eval.
UNCERTAINTY DUE TO ASSUMING A PLANE PARALLEL SINGLE LAYER HOMOGENEOUS CLOUD IN FORWARD MODEL (PPSH)						
1	Vertical variability of cloud properties	This is a significant source of uncertainty, particularly for the heritage channel retrieval.	Poulsen et al. (2012) Siddans et al. (2010) Cloud Model study	Multi layer cloud, Vertical size distribution, Mixed phase cloud, Thick aerosol over cloud, PSCs	SYS	SIM
1	Application of plane parallel model to a inhomogeneous cloud field		Heidinger and Stephens (1998) Siddans et al. (2010)	Sub pixel cloud, Shadowing, Viewing cloud sides, Photon source/loss around cloud edges	SYS	SIM
2	Uncertainty in assuming infinitely thin cloud				SYS	SIM

UNCERTAINTY IN CLOUD IDENTIFICATION						
2	Error due to partly cloudy/clear (aerosol)pixel			No aerosol assumed in clear fraction, Misidentification of aerosol as cloud	SYS	SIM
1	Error due to misidentifying cloud/aerosol in cloud mask	Very significant source of uncertainty.				
1	Error due to misidentifying snow and sea ice as cloud			Usually too much cloud detected over snow/sea ice	SYS	
1	Cloud mask varies between day/night					
UNCERTAINTY IN FORWARD MODEL						
3	Trace gas errors H2O and O3	.1-2% uncertainty in TOA radiances	Siddans et al. (2010) Cloud model for Operational retrievals EUMETSAT report	Most significant for low cloud and 0.87um channel	SYS RAN	FM
4	Trace gas CFC errors			Only applicable to IR channel	SYS RAN	

3	Error in modelling transmission given 'true' profiles	<1% for SEVIRI Uncert. in transmis.	Siddans et al. (2010) Fast transmittance modelling of MSG and MTG solar channels for cloud retrievals.	Highest for 0.87, low clouds and large LZA.	SYS RAN	FM
3	Gaseous absorption approximations	0.2% Uncert. in transmis	Kratz (1995)		SYS RAN	
1	Cloud optical properties (extinction coefficient, SSA and phase function.) Particle size distribution.	Small for water Significant for ice Important for aerosol	Cooper et al. (2006) give values for MODIS of 20-30% uncertainty in cloud optical depths	Could be performed via a comparison of scatter phase functions. More significant for vis than IR. IR channels more useful for thin cloud	SYS	FM SIM
4	Error assuming a single central wavelength when calculating LUTS.	small		Could be assessed by calculating LUTS for higher resolution wavelength.	SYS	
4	Uncertainty in DISORT			Number of streams used.	SYS	??
3	Forward Model error e.g RTTOV		Saunders et al. [R-13]	Comparisons with LBL models	SYS	SIM

UNCERTAINTY IN AUXILARY DATA SETS						
2	Uncertainty in NWP T- profile q-profile			T and q vertical profile uncertainty more important for IR than visible channels. Look at B-matrix	RAN SYS	
3	T-surface	1/5K sea/land		Comparison with buoy over sea. Likely to be significant over strongly diurnal varying surfaces such as deserts.	SYS RAN	
	Error in Sea ice and snow map	?		Potential to create a discontinuity if different seaice/snow information used for different time periods	SYS	
2	Surface Reflectivity/BRDFError	MODIS land albedo accuracy .02-.05	Liu et al. (2009) MODIS albedo accuracy paper	Significant over some land surfaces such as ice and desert- correlated???	RAN	
2	Emissivity error	5% over deserts <1% over sea uncertainty in emissivity value		Significant over land surfaces in particular deserts, small over sea. PCA may need to be applied.	RAN	FM

2	Horizontal representivity of axillary data sets			Snow, ice, brdf,emissivity	
2	Horizontal represent. of NWP parameters			T,P,q,uv	
ADDITIONAL SOURCES OF UNCERTAINTY					
4	Uncertainty in reference solar spectra	2-3% in TOA reflect. Only sig. for thin cloud.	Sayer et al. (2010)	Not significant for AATSR; significant for AVHRR	SYS
4	Uncertainty in surface heterogeneity	Small negligible	Lyapustin et al.	Shadowing from surface	
4	Sphericity of the atmosphere	negligible			SYS
1	Uncertainty in Diurnal correction				
2	Effects due instrument geometry			Uncertainty usually increases to edge of swath	
2	Effects of spatial and/temporal averaging to L3				
3	Polarisation			Will effect atmospheric RTand possibly surface reflection	

INSTRUMENT UNCERTAINTY						
1	Coregistration error between channels			Could be more significant for AATSR when MERIS is used (MERIS push broom AATSR conical scanner coregistration n will vary approx. 10% with orbit height as well)	SYS BIAS	FM
1	Calibration uncertainty	3% uncertainty on TOA radiance cf reference	Prelaunch calibration report (AATSR)		RAN	FM
1	Calibration offset	TBD in CCI	Found via vicarious calibration or SNO corrected before processing.	Offset compared with MERIS/MODIS	SYS BIAS	
3	Geolocation error			Land/sea mismatch	SYS?	FM
3	Spectral shift	0.2K 12um (AATSR)		AATSR only IR channels?	SYS	
2	Channel Degradation			Could be quantified in the above.	SYS	FM

Table 3-2 *Sources of Uncertainty in CC4CL retrievals. FM: Forward model error, APE: Apriori error, SIM: Assess error via simulations or other means and report separately, RAN: Random error, SYS: Systematic error, Priority 1(Most important - essential to quantify) to 4(Least important - only quantify if time)*

4. Methodology to retrieve and evaluate uncertainties

A number of methodologies are employed to retrieve and evaluate the uncertainties on the Cloud_cci products.

a. The uncertainties are propagated through the model and are provided with the retrieved parameters.

b. The uncertainties are evaluated by performing simulations and sensitivity studies which are reported or referenced.

c. The uncertainties are evaluated by compositing.

d. Stability: A method to evaluate the stability of the cloud parameters has yet to be proposed.

Each of these methodologies is outlined in more detail below.

4.1 Uncertainties propagated via the model:

Uncertainty characterisation is a feature of the CC4CL algorithm. Wherever possible, uncertainties are propagated from the initial sources throughout the retrieval to the final retrieved product in a rigorous mathematical format.

The CC4CL algorithm uses the optimal estimation (OE) method described in [R-9] to optimise the retrieval and characterise the uncertainties. OE assumes there is a forward model (FM), which in this case is a single layer, to simulate measurements:

$$\dot{\mathrm{y}} \;=\; F(\hat{\mathrm{x}}) + \epsilon$$

Equation 1: FM

where ***y*** is a vector comprising all relevant measurements made by the sensor and ***x*** is the "state" vector containing parameters to be optimised by "inverting" the forward model, *F*. ε is the measurement error, described by the covariance S_y.

OE makes use of *a priori* knowledge to arrive at a statistically optimal solution to the generally otherwise under-constrained inverse problem by minimising the "cost function":

$$\chi^2 \;=\; (\dot{\vec{y}} - F(\hat{\vec{x}}))^T \mathrm{S}_y^{-1} (\dot{\vec{y}} - F(\hat{\vec{x}})) + (\vec{x} - \vec{a})^T \mathrm{S}_a^{-1} (\vec{x} - \vec{a})$$

Equation 2: Cost

where ***a*** is the *a priori* estimate of the state vector, which has covariance about the true state S_a.

The solution for *x* which minimises the cost function is found by Marquardt-Levenberg iteration:

$$\vec{x}_{i+1} = \vec{x}_i + (\mathrm{S}_a^{-1} + \mathbf{K}^T \mathrm{S}_y^{-1} \mathbf{K} + \gamma \mathbf{I})^{-1} [\mathbf{K}^T \mathrm{S}_y^{-1} (\dot{\vec{y}} - F(\vec{x}_i)) - \mathrm{S}_a^{-1} (\vec{x}_i - \vec{a})]$$

Equation 3: Marquardt-Levenberg iteration

where **K** is the weighting function matrix comprised of the derivatives of the measurements with respect to each element of the state vector, calculated by the FM, and γ is a scaling constant.

Random errors on the retrieved values are described by the solution covariance:

$$\mathbf{S}_x = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K})^{-1}$$

Equation 4: Estimated retrieval covariance

The sensitivity of the retrieval to perturbations in measurements is given by the retrieval contribution function or gain matrix:

$$\mathbf{D}_y = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K})^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}$$

Equation 5: Measurement contribution function

Multiplying the measurement contribution by the weighting function matrix gives the averaging kernel, which characterises the sensitivity of the retrieval to perturbations in the true state:

$$\mathbf{A} = \mathbf{D}_y\, \mathbf{K}$$

Equation 6: Covariance of x from measurement error covariance

The trace of the averaging kernel gives the degrees of freedom for signal, a measure of how many independent parameters can be extracted from the measurements. Information content is another useful feature which quantifies the number of bits of information added by measurements:

$$H = \frac{1}{2}(\log_2(det\mathbf{S}_a) - \log_2(det\mathbf{S}_x))$$

Equation 7: Information content

The error on ***x*** from an error source, denoted by subscript b, not accounted for in the measurement covariance matrix is,

$$\Delta\hat{\mathbf{x}}_b = \mathbf{D}_y\mathbf{K}_b\Delta\mathbf{b}$$

Equation 8: Error on x due to parameter error

For an error in some model parameter, Δb, assuming $\mathbf{K}_b$ is the sensitivity of the FM to that parameter. Where the error source can be quantified by a covariance matrix, the resulting covariance in **x** is

$$\mathbf{S}_{x:b} = (\mathbf{D}_y\mathbf{K}_b)\mathbf{S}_b(\mathbf{D}_y\mathbf{K}_b)^t$$

Equation 9: Covariance of x from parameter covariance

4.2 Uncertainties evaluated by simulations:

There exist some sources of uncertainty that cannot be easily propagated through the algorithm. These may neither be easily quantifiable nor Gaussian by nature. Systematic uncertainties are usually of the latter kind, an example being the uncertainty incurred by using a single layer model to retrieve multi-layer clouds. In such a case, the uncertainty can be evaluated using simulated data. Simulations of this type have been performed for CC4CL in [R-18] to evaluate the performance of the algorithm for very thin cloud and multi-layer scenarios. In [R-18] Validation was undertaken with a reference forward model, i.e. a more extensive forward model that attempts to eliminate some of the most important assumptions in the “fast” solution. Results show that, in relation to the simple scalar operators, for optical thicknesses greater than 10, the errors are comparable to instrument noise, but it should be noted that this error is the difference between the reference forward model and the “fast” forward model and not a measure of the total errors in forward modelling. At small optical thicknesses (less than 0.1–1.0) the errors become larger, especially at optical thicknesses approaching the critical regime of unity, where the contribution of single and multiple scattering to the total shortwave signal are comparable

4.3 Uncertainties evaluated by compositing:

We can easily describe several sources of uncertainty, e.g. view angle bias, surface type heterogeneity influences, etc., through evaluation of daily and monthly averages applying different filtering criteria. Such methods have been performed successfully for MODIS [R-8].

4.4 Uncertainties in statistical models:

For some cloud products obtained by training statistical models it is also difficult to propagate the uncertainty through the algorithm. For example, the cloud mask consists of binary values that are provided by an artificial neural network (ANN). In particular, the cloud mask is obtained by combining a set of neural networks that have been optimized for different situations (i.e. a representative selection of globally and seasonally resolved imager data) and trained both for day and night. In this case, the ensemble outputs can be used to provide an uncertainty estimate per sample: a low variance of the ensemble of ANNs indicates high confidence on the accuracy of the output, while an average output activation distant from the decision boundary indicates high confidence on the class assignment. Therefore, the coefficient of variation of the ensemble outputs is selected to estimate the uncertainty of each detection.

4.5 Stability

Last but not least, one can assess and validate the stability of long-term data sets in order to derive the uncertainty of the retrieved parameters. The characterization of their climatological stability is an important requirement for applying them in climate variability and trend analyses. Discontinuous temporal sampling and satellite drift (if occurring) have to be corrected by developing and applying methods to statistically reconstruct a continuous diurnal cycle of the parameters of interest. This will in turn allow producing a temporally and spatially consistent ECV suitable for climate monitoring.

In Cloud_cci phase 2 there is a work package, which characterizes the stability (i.e. homogeneity) of the retrieved cloud products by means of:

- Evaluation of climatological stability and homogeneity of the AVHRR and MODIS cloud cover over Central Europe by use of SYNOP on the monthly and yearly time scale.
- Analysis of diurnal cycle of the AVHRR and MODIS cloud cover by use of a Meteosat-based cloud climatology (CM SAF) as well as SYNOP data over Central Europe.

5. Evaluation of uncertainties in the products

A number of techniques can be used to evaluate the uncertainties in the products. Multiple techniques are necessary as they address different aspects and sources of uncertainty. No single technique can evaluate all the uncertainties on cloud parameters.

a. Comparison of retrieved products with collocated, independent data sets.

b. Evaluation of product uncertainties using collocated data sets.

The first technique applies the tools and techniques developed within the Cloud_cci team for the round robin data sets, as documented in the PVSAR report. These tools compared MODIS/AVHRR data to temporally and spatially collocated Calipso, Cloudsat and AMSRE data. The Calipso and Cloudsat cloud data were used to evaluate the cloud top height, phase, and cloud mask. The AMSRE data evaluated the LWP, which is an indirect validation of the combination of optical depth and effective radius.

Examples of the comparisons undertaken are shown in Figure 5-1 and Figure 5-2. This method is particularly useful for evaluating the systematic biases and uncertainties due to assuming a PPSH model for the cloud retrievals. In this kind of comparison, CALIPSO data help to distinguish different cloud scenes: clouds with COD > 3, single layer clouds with COD < 3, clouds with COD < 3 and lower clouds underneath.

Currently, there is no high accuracy data set available that would facilitate the independent evaluation of optical depth and effective radius. The Cloud_cci team will pursue and review any useful validation data sets as it becomes aware of them in order to address this deficiency.

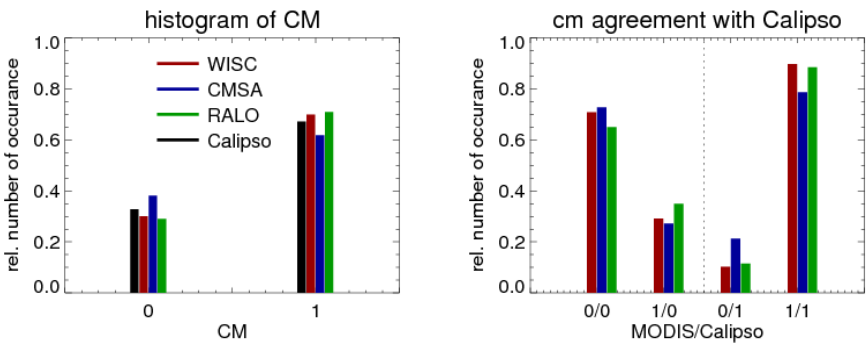

Figure 5-1 *Comparison of MODIS cloud mask algorithms from CM SAF, WISC and RALOX algorithm with Calipso Cloud mask from the Product Validation and Algorithm Selection Report [R-7].*

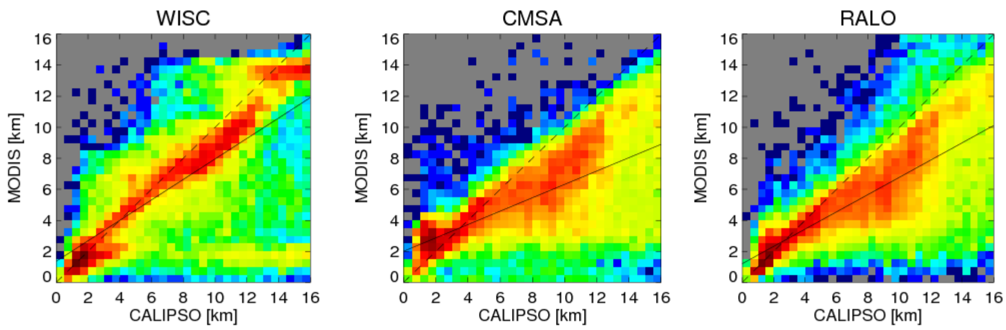

Figure 5-2 *Comparison of MODIS Cloud top height retrievals from different algorithms with CALIPSO cloud height.*

A second technique for validating the random product uncertainty is to compare the uncertainty of the retrieval (binned from low to high values) vs. the similarly binned RMS of the difference between a 'truth' value and the collocated retrieved value. This method is particularly appropriate for evaluating random uncertainties such as instrument measurement uncertainty and other uncertainties that are propagated through the retrieval, e.g. uncertainties of surface reflectance and emissivity. The technique is limited to validating cloud parameters for which accurate independent temporally and spatially collocated data sets exist.

Variable	Sources of validation/Comparison Data	Accuracy	Comment/reference
CTH	Calipso	Sensitive to the top of the cloud- accurate	In the case of CTH it is necessary to consider validating the 'effective' cloud top height. This will take into account the penetration depth of the satellite channels.
CTH	Cloudsat	Sensitive to cloud profile	As above
Optical depth	N/A		Comparison with MODIS
Effective radius	N/A		Comparison with MODIS
Ice water path	Calipso DARDAR product	?	Only possible for single layer thin ice cloud
Liquid water path	AMSRE	?	Care has to be taken to validate clouds only in the scenarios where AMSRE is the most accurate.
Cloud fraction	Calipso	?	
Cloud fraction	Synop		Representivity Issues
Cloud albedo	CERES	?	

Table 5-1 *Sources of data for validation and Comparison with CC4CL cloud products*.

In both cases it is useful to filter (using the retrieval cost) the cloud data sets for cloud scenarios that are well-described by the PPSH scenario.

An inherent problem in this technique is that it requires a good statistical sample of collocated data covering the spectrum of cloud properties. This is possible for the MODIS retrievals but Calipso/Cloudsat overlap the Envisat paths only at the poles. This restricts the comparison to a specific meteorological region, for example tropical convective clouds uncertainties cannot be validated for AATSR/MERIS. However, for lack of other good validation data it may be sensible to identify a year's (TBD) worth of collocations with Calipso in this region and perform the validation on these scenes.

6. Uncertainty characterisation of Level 3 products

6.1 Current implementation

The current knowledge of uncertainty in existing L3 cloud retrievals and evaluation is currently summarized in the GEWEX Cloud Assessment report [R-11].

A very useful report on error propagation when generating gridded satellite data sets has been produced by Ralf Bennartz [R-12], which informs the discussion wrt. correlated uncertainty below.

Level 3 products will utilize the provided L2 uncertainty information as far as possible. Of several plausible, Table 6-1 lists approaches the uncertainty information currently produced as part of a level 3 product.

Table 6-1 Uncertainty terms included in version 2.0 of Cloud_cci data products.

Uncertainty Variable Name in file and code	Descriptive name	Description/Comments
Variable_std	Standard deviation σ_{std}	See item (1) below.
Variable_unc	Mean uncertainty $\langle\sigma_i\rangle$	See item (2) below.
Variable_prop_unc	Propagated uncertainty σ_{prob}	See item (3) below.
Variable_corr_unc	Correlated uncertainty σ_{corr}	See item (4) below.

1. The standard deviation S_{STD} of x is calculated by:

$$\sigma_{\text{std}}^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(x_i - \langle x\rangle)^2 \qquad \textbf{Equation (6-1)}$$

2. The mean uncertainty is calculated by:

$$\langle\sigma_i\rangle = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\sigma_i) \qquad \textbf{Equation (6-2)}$$

3. Propagated uncertainty is calculated by:

$$\sigma_{\text{prob}} = \frac{1}{\mathrm{N}}\sqrt{\sum_{i=1}^{\mathrm{N}}(\sigma_i^2)} = \sqrt{\frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}}(\sigma_i^2)} = \sqrt{\frac{1}{\mathrm{N}}\langle\sigma_i^2\rangle} \qquad \textbf{Equation (6-3)}$$

4. Correlated uncertainty is calculated by:

$$\sigma_{\text{corr}} = \sqrt{\sigma_{\text{std}}^2 - (1-c)\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\sigma_i^2) - (\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\sigma_i))^2}} = \sqrt{\sigma_{\text{std}}^2 - (1-c)\sqrt{\langle\sigma_i^2\rangle - \langle\sigma_i\rangle^2}} \qquad \textbf{Equation (6-4)}$$

c denotes the uncertainty correlation.

For variables which are part of the state vector (CTP, COT, CER) or are directly inferred from them (CWP, LWP, IWP, CLA) the above mentioned Level-3 uncertainty terms can be used to approximate two important monthly properties: the natural variability of the cloud variables observed ($\sigma_{\mathrm{natural}}$) and the uncertainty of their monthly mean ($\sigma_{\langle x \rangle}$). Following Bennartz et al. (R-12) and Stengel et al. (2017) these two terms can be calculated by:

$$\sigma_{natural}^2 = \sigma_{std}^2 - (1-c)\langle \sigma_i^2 \rangle \qquad \textbf{Equation (6-5)}$$

And

$$\sigma_{\langle x \rangle}^2 = \frac{1}{N}\sigma_{natural}^2 + c\langle \sigma_i \rangle^2 + (1-c)\frac{1}{N}\langle \sigma_i^2 \rangle \qquad \textbf{Equation (6-6)}$$

The terms σ_{std} and $\langle \sigma_i \rangle$ are directly stored in the Level-3 files. The mean of the squared uncertainties $\langle \sigma_i^2 \rangle$ can be inferred from the stored σ_{prob} fields by rearranging Equation 6-3.

Figure 6-1 shows most important terms discussed above using the example of cloud optical thickness in 2008/06. For more information about this example see Stengel et al. (2017).

6.2 Future uncertainty characterization

It seems more practical and user friendly to store $\sigma_{\mathrm{natural}}$ and $\sigma_{\langle x \rangle}$ in the output files (using a predetermined correlation). In addition, to enhance flexibility, it seems also desirable to store $\langle \sigma_i^2 \rangle$. Together with $\langle \sigma_i \rangle$, this would enable a straight forward recalculation of $\sigma_{\mathrm{natural}}$ and $\sigma_{\langle x \rangle}$ for any given correlation.

- Probability of ML cloud calculated using the Pavolonis overlap flag

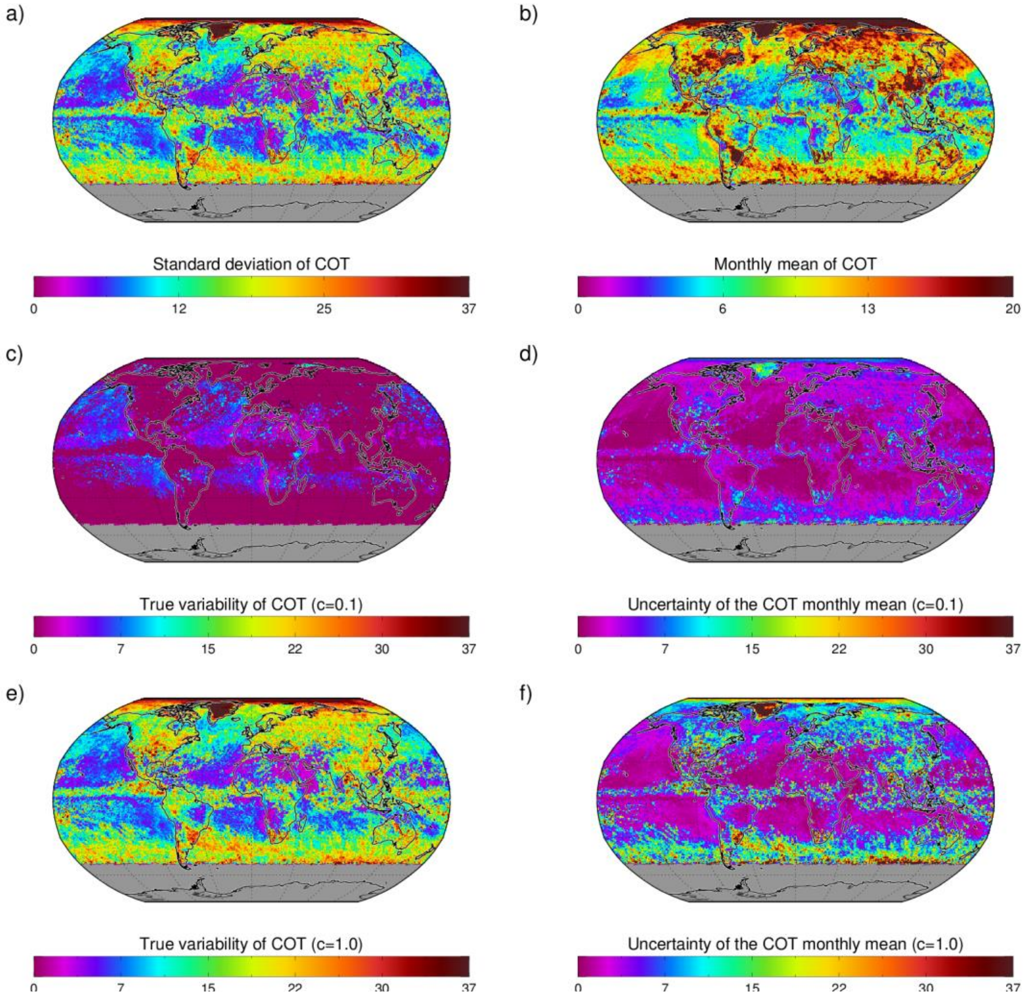

Figure 6-1 Monthly standard deviation (a) and monthly mean (b) for cloud optical thickness (COT). Panels (c) and (d) show the estimated natural variability and uncertainty of the mean (d) for a correlation of 0.1. Panel (e) and (f) are as panels (c) and (d) but for an uncertainty correlation of 1.0. All data is from AVHRR-PM in 2008/06. Figure taken from Stengel et al. (2017).

7. Cloud parameter uncertainty validation

This section demonstrates some of the analyses that have been performed to validate the uncertainty estimate provided in the Cloud_cci products. Each cloud parameter has an uncertainty estimate that is the result of measurement and forward model (surface and cloud inhomogeneity)uncertainty propagated through the retrieval (see equation 4 for details on the solution covariance).

7.1 CTH uncertainty validation

7.1.1 CC4CL

A technique to validate the uncertainty is to compare the retrieved uncertainty with the uncertainty derived from the difference of an accurate collocated data source with the retrieved products. The only retrieved cloud property for which we have accurate validation data is the CTH which can be derived from collocated Calipso and/or Cloudsat observations.

5 days of data of collocated MODIS and Calipso retrievals were analysed. Figure 7-1 shows a plot of the MODIS/Heritage – Calipso CTH difference divided by the retrieved uncertainty of the MODIS Cloud top height product on the left, on the right the effective cloud top height is calculated by using the Calipso optical depth profile to estimate the height one optical depth into the cloud.

If the uncertainty is assumed to be random then approx. 66% (2 sigma) of the results should lie within ± 1 Values less than 1 indicate the retrieved uncertainty is too high while values greater than 1 indicate the uncertainty is too low.

This systematic uncertainty is not propagated into the uncertainty reported in the product as the OE framework enables only Gaussian like uncertainty to be propagated. Systematic uncertainties in principle should be modelled and the bias removed however in the case of cloud retrievals it is almost impossible to predict the systematic effect of missing thin upper layers or the effects of multiple layers of cloud. This has been discussed in Poulsen et al (2012) and Stengel et al (2015).

Figure 7-1 shows a comparison of the heritage single layer retrieval and a single layer retrieval that uses multiple MODIS channels (0.67, 0.87, 1.2, 2.13, 3.7, 6.7, 7.33, 11, 12, 13.3, 13.6, 13.9, 14.2 µm). The heritage retrieval uncertainty significantly underestimates the real uncertainty. This underestimation in agreement to a more comprehensive study documented in Section 6 of PVIR (2018). Furthermore, a large positive bias can be seen where the heritage CTH is significantly lower than the collocated Calipso CTH. This difference is caused by the inability of the forward model to retrieve the CTH of predominantly multi-layer cloud (usually thin cirrus over a liquid cloud layer). This is a systematic uncertainty which is not accounted for in the forward model. It is important when evaluating the quality of a cloud retrieval that the retrieved uncertainty is not considered in isolation to the Cost. The Cost is a good indicator of how well the measurements fit the model. Cloud retrievals which do not fit the single layer model often have a high cost. Reducing the number of collocations to those that satisfy a cost threshold e.g. cost less than 5 results in significantly more retrievals with accurate uncertainty estimates.

The additional channels particularly the 7 µm and 13 and 14 µm channels in the multi-channel retrieval increase the sensitivity to the vertical profile and thin ice layers. The systematic bias that was previously observed with the heritage channel retrieval is much reduced. Figure 7-1 also evaluates the associated retrieval uncertainty. The lower plots show the ratio of CTH error to retrieval uncertainty where CTH error is defined as the difference between the Calipso CTH and the retrieved cloud top height. The distribution of observations within the Gaussian fit is much improved for the case of the multi-channel retrieval. The systematic bias in the heritage channel retrieval has been reduced.

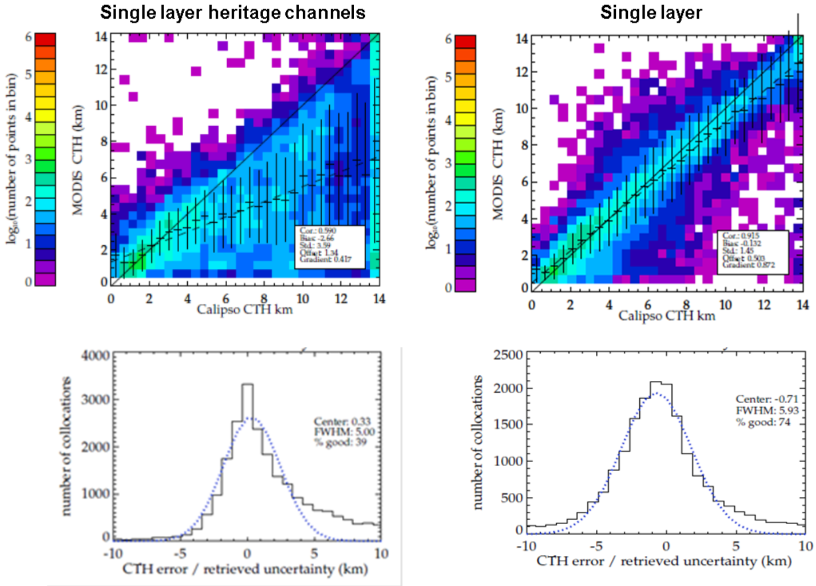

Figure 7-1 *Shows a comparison of the upper layer cloud top height with Calipso (top) and the corresponding uncertainty validation (bottom) (Calipso - CC4CL MODIS)/retrieved uncertainty.. The two retrievals shown are the single layer retrieval algorithm for the heritage CC4CL retrieval which uses only 5 channels (left), and a multi-channel single layer retrieval from CC4CL (right).*

7.1.2 FAME-C

The FAME-C CTP random uncertainties are directly derived from the input uncertainties of the optimal estimation. For this evaluation the CALIOP 1kmCLay Product, with an assumed random uncertainty of 10hPa, was used. The AATSR CTP of FAME-C mostly has an uncertainty in the same order of magnitude, while the MERIS CTP uncertainty is either around 10hPa or in the order of 500hPa. For the remaining part of this analysis the biases of CTP_{MERIS} and CTP_{AATSR} to CTP_{CALIOP} were removed when the uncertainty is validated.

Comparison the difference of FAME-C CTP and CALIOP ($|CTP_F$-$CTP_C|$) to the total uncertainty ($\sqrt{unc_F^2+unc_C^2}$) is shown in the scatter plots (Figure 7-2). In 85.6% of all investigated retrievals the difference between CALIOP and AATSR CTP is larger than the combined uncertainty, while for 14.4% the difference is smaller than the uncertainty. For MERIS CTP, 35.2% percent of all retrievals have a deviation from CALIOP larger than the uncertainty. For 64.8% of the cases the uncertainty is larger than the deviation to CALIOP. The shape of the scatter does not indicate a correlation between the amplitudes of uncertainties and deviations.

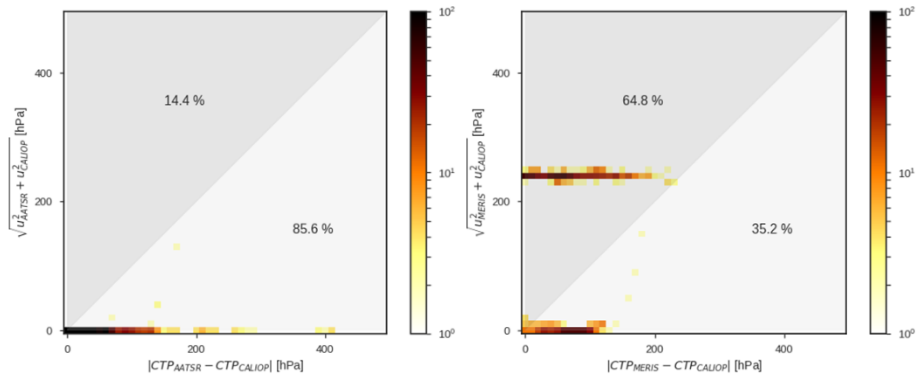

Figure 7-2 *Two dimensional frequency histograms of FAME-C CTP_{AATSR} (left) and FAME-C CTP_{MERIS} (right) deviations from CALIOP (x-axes) compared to the combined uncertainties (y-axes). Also given are the relative numbers of cases for which the deviation is smaller than the combined uncertainty.*

Figure 7-3 shows an example cross section (case study 1) reporting the CTP retrievals for CTP_{MERIS}, CTP_{AATSR} and CTP_{CALIOP}. It seems clear that FAME-C CTP_{AATSR} is closer to the CTP_{CALIOP} measurement than CTP_{MERIS}. Figure 7-4 shows the same data as in Figure 7-3 after removing the bias to CALIOP. In addition, Figure 7-4 shows comparisons of corresponding deviations to the combined uncertainties. As mentioned above FAME-C MERIS uncertainties are much larger than for AATSR. For AATSR the uncertainties are much smaller than the CTP deviation to CALIOP for most pixels. For MERIS the uncertainties are much larger than the deviation to CALIOP for most pixels. For case study 2 (Figure 7-5 and Figure 7-6, a similar result is found as in case1, although the mismatch in amplitude between the uncertainties and the CTP deviation to CALIOP seems smaller.

In can be concluded that further work has to go into the CTP retrievals and CTP uncertainties. The uncertainties are generally too small and show low correlation to the actual deviation from the truth (CALIOP here). Earlier studies have shown that the CTP retrievals can be as good as 25hPa (Fischer and Kollewe, 1994; O'Brien and Mitchell, 1992, Preusker and Lindstrot, 2009). Future version of FAME-C should be able to reach this accuracy and to provide uncertainties that properly estimate the random uncertainty.

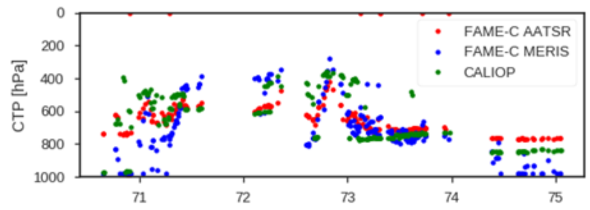

Figure 7-3 *Case study 1. Comparison of CALIOP CTP with FAME-C CTP_{MERIS} and FAME-C CTP_{AATSR}.*

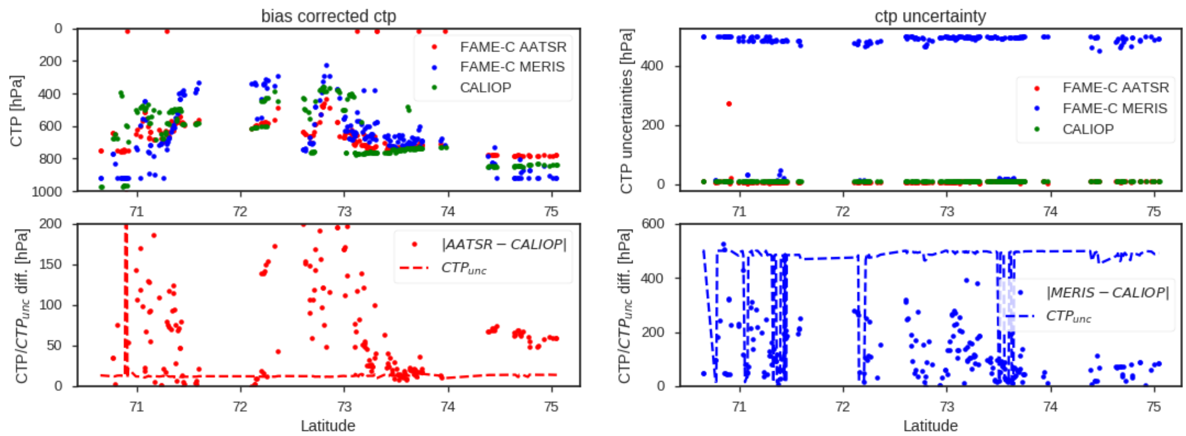

Figure 7-4 *Case study 1 continued. Top-left: as Figure 7-3 but after bias removal. Top-right: CTP uncertainties; bottom-left: absolute difference of CTP_{AATSR}-CTP_{CALIOP} (after bias removal) and $CTP_{AATSR,unc}$; bottom-right: absolute difference of CTP_{MERIS}-CTP_{CALIOP} (after bias removal) and $CTP_{MERIS,unc}$.*

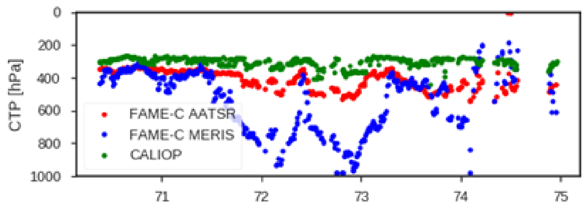

Figure 7-5 *Case study 2. Comparison of CALIOP CTP with FAME-C CTP_{MERIS} and FAME-C CTP_{AATSR}.*

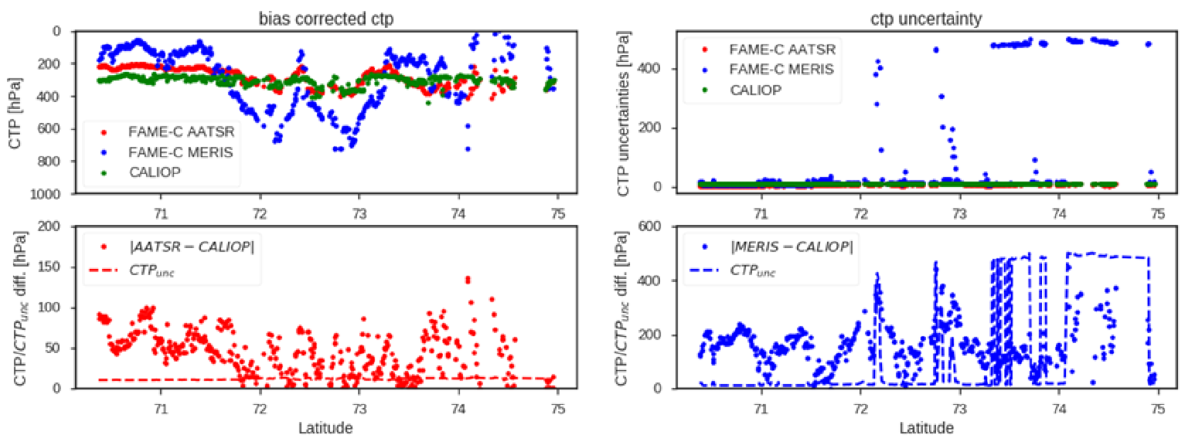

Figure 7-6 *Case study 2 continued. Top-left: as Figure 7-5 but after bias removal. Top-right: CTP uncertainties; bottom-left: absolute difference of CTP_{AATSR}-CTP_{CALIOP} (after bias removal) and $CTP_{AATSR,unc}$; bottom-right: absolute difference of CTP_{MERIS}-CTP_{CALIOP} (after bias removal) and $CTP_{MERIS,unc}$.*

7.2 Uncertainty in multi-layer cloud conditions

The AVHRR heritage channel retrieval obtains most of its information on cloud top height from the 10 and 11 µm channels. These channels are sensitive to the temperature of the cloud. In order to retrieve an accurate cloud top height this technique assumes the upper layer is optically thick, with no contribution from the surface or lower cloud layers. CC4CL includes the surface temperature in the state vector and the visible channels provide some additional information on the transparency of the cloud however it does not account for multiple layers of cloud. When multiple layers of cloud are present and the upper cloud layer is thin then the CTH retrieved is an effective cloud top height located between the lower and upper layers. This effective CTH is illustrated in the top plot of **Figure** 7-7 where the green, blue and red points illustrate the CC4CL retrieved height, green indicates a low cost, red a high cost. Note where the upper layer is thick, no signal from the lower layer is observed and the retrieved height of the upper layer is accurate. Where the upper layer is thin the retrieved height is the 'effective cloud top height'. In most of these cases the cost will be high as the observed radiances do not fit a single layer cloud model. The retrieved cloud effective radius in this case is similarly a weighted effective radius between the lower, usually liquid layer and the upper, usually ice, layer, see the Multi-layer validation report for more explanation and examples. This uncertainty is not accounted for in the retrieval as it is a systematic uncertainty that is highly variable. While the L2 CC4CL product includes the cost value which will inform the user of the quality of the product, the L3 product does not have this information to indicate regions where this uncertainty is likely to significantly affect the accuracy of the product. How to communicate this uncertainty should be considered in future releases of cloud products, particularly those relying on the heritage channel selection. One metric to consider might be percentage of retrievals in a grid box with low cost.

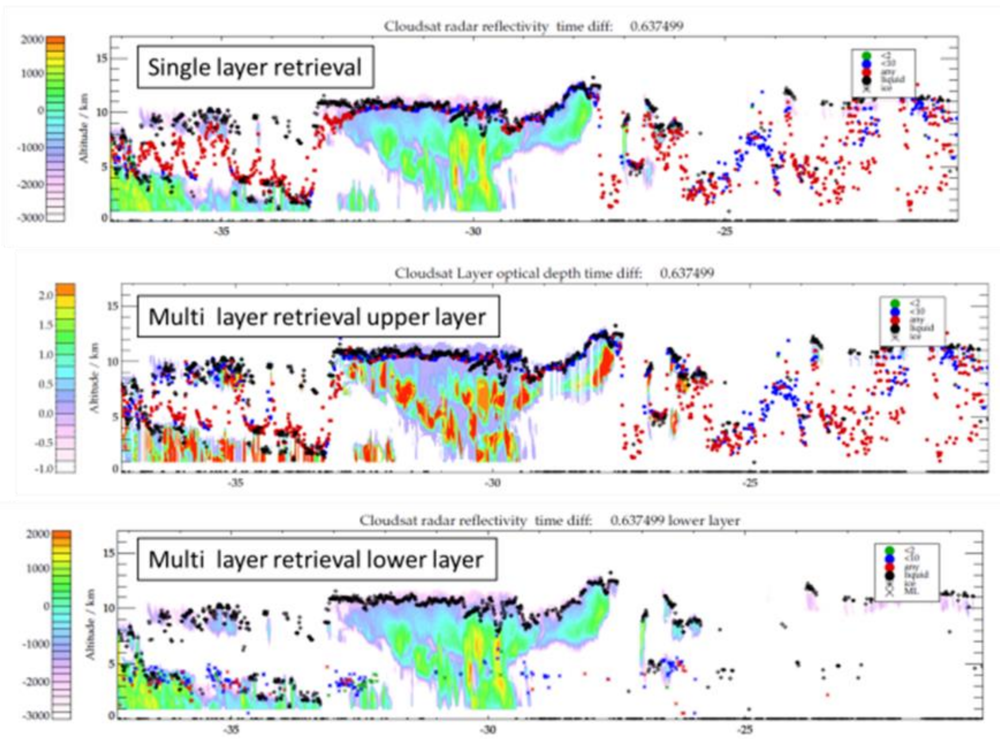

Figure 7-7 *Shows a cross section of a MODIS multi-layer retrieval (middle and lower panel) from 06/20/2008 compared with single layer retrieval (top panel). Collocated with coincident Cloudsat data. Red, blue green dots show the retrieved cloud top height, green dots have a low cost, red dots a high cost. Black dots show the actual cloud top height.*

8. Documentation of uncertainties in the products

Short name	Longname	units
CTH_uncertainty	Standard_uncertainty of_CTH	km
CTP_uncertainty	Standard_uncertainty of_CTP	hPa
COT_uncertainty	Standard_uncertainty of_COT	
CER_uncertainty	Standard_uncertainty of_CER	µm
COST	i.e. the measure of how well the measurements fit the PPSH model used.	
CC_TOTAL_uncertainty	Standard_uncertainty_of_CC_TOTAL	%
CTT_uncertainty	Standard_uncertainty_of_CTT	K
CWP_uncertainty	Standard_uncertainty_of_CWP	g/m^2
LWP_uncertainty	Standard_uncertainty_of_LWP	g/m^2
IWP_uncertainty	Standard_uncertainty_of_IWP	g/m^2
Cloud_Albedo_uncertainty	Standard_uncertainty_of_Cloud_Albedo	

Table 8-1 *Description of variables used to describe uncertainty in the CC4CL products.*

9. Conclusions

A document has been produced that outlines cloud product uncertainty sources, quantification methodology, and some initial evaluation. The cloud parameters are subject to a number of sources of uncertainty. Future effort will be put into identifying the most significant sources of uncertainty and either characterising the effect of the uncertainty on the interpretability of the product or propagating the sources of uncertainty though the retrieval in order to retrieve realistic uncertainty estimates.

The CTH uncertainty in the CC4CL products have been evaluated with active sensors. For single layer clouds the retrieved uncertainty is a reasonable estimate of the uncertainty except for the cases of thin cirrus over a liquid cloud layer. Users are recommended to consider both cost and retrieval uncertainty when evaluating cloud top height products.

Appendix 1 - ESA Climate Change Initiative Guidelines on Uncertainty Characterization

Version 2.0; 2 Nov 2010

Introduction

During the first CCI Colocation Meeting at ESRIN on 15-17 September 2010 an open discussion on uncertainty characterization was held, attended by members of all CCI projects. Subsequently, a "drafting team" was tasked to discuss common issues relevant to uncertainty characterization and to draft relevant useful guidelines, including a common table of contents for the CCI "Error Characterization" document deliverable. The initial conclusions of the drafting team were presented during the final plenary session of the colocation meeting. These conclusions, and in particular the definitions were further worked on during the following few weeks, resulting in this document. The drafting team consisted of the following members:

ESA	Simon Pinnock
CCI Cloud & Aerosol	Don Grainger
CCI Fire	Martin Schultz
CCI Land Cover	Pierre Defourny
CCI SST	Chris Merchant
CCI Ocean Colour	Roland Doerffer
	Sylvia Kloster
CCI CMUG	David Tan

Presentations from the colocation sessions and plenary are available on the web at: http://earth.eo.esa.int/workshops/esa_cci/cci_coloc.html

The main conclusions of the Uncertainty Characterization drafting team were:

1. CCI projects should use a common definition of terms in their Uncertainty Characterization work.
2. A common table of contents for the "Error Characterization" document deliverable was proposed

These conclusions were translated into the following two proposed guidelines.

Guideline #1 - Common Definition of Terms

(a) The CCI deliverable "Error Characterization Document" should be renamed "Uncertainty Characterization Document" as this would be better aligned with current usage of the terms "error" and "uncertainty".

(b) As the uncertainty characterization document will be read by users who are likely to use outputs from several CCI projects, it is desirable that all CCI projects use the same definition of terms.

The following example text could be used as Section 2 of the 'Uncertainty Characterization Document'. If individual CCI project teams chose to provide their own Section 2 they should not modify the given definitions.

Definition of Terms

Describing error and uncertainty

A measurement is a set of operations having the object of determining the value of a quantity. Following BIPM (2008) it is helpful to define the term measurand as

> **measurand:** particular quantity subject to measurement,

so that the phrases 'true value of a quantity' and value of the measurand are synonymous.

Very few instruments directly measure the measurand. Generally an instrument reports the effect of a quantity from which the magnitude of the measurand is estimated. As an example, an instrument sensitive to infrared light might be used to measure the temperature of an object.

The process of measurement is inexact, so that difference between a measured value and the value of the measurand is called the error. Traditionally (e.g. Beers, 1975) the word 'error' has also meant a numerical value that estimates the variability of the error if a measurement is repeated (i.e. a width of the distribution of possible errors). This dual meaning of "error" can lead to confusion or ambiguity. To separate these meanings and avoid confusion the BIPM (2008) definitions are used, i.e.

> **error (of measurement)**: result of a measurement minus a true value of the measurand,
>
> **uncertainty (of measurement)**: is a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

Except in a few cases the "true" value of the error is not known, and the magnitude of the error is hypothetical. An error is viewed as having a random component and a systematic component. Following BIPM (2008) the definitions of these terms are:

> **random error:** result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions,
>
> **systematic error:** mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions minus the true value of the measurand.

In general terms the random error is variable from measurement to measurement, whereas the systematic error is the same for each measurement. Although it is not possible to compensate for the random error, its effect on uncertainty in our estimate of the measurand can usually be reduced by averaging over a number of independent repeat observations.

The statistical distribution of random error can be described by a probability density function (pdf) of which the **expected value** (i.e., the average over the pdf) is zero. As the random error often arises from the addition of many effects the central limit theorem suggests that a Gaussian distribution is a good representation of this pdf. Therefore the random uncertainty value commonly adopted for a single observation is equal to the one-sigma standard deviation that would be obtained from repeated measurements of the same quantity under the same conditions. If N repeated uncorrelated observations are available, the random uncertainty is the one-sigma standard deviation multiplied by a factor of $1/\sqrt{N}$ (under the Gaussian assumption). The smallest possible change in value that can be observed can be taken as ½ the uncertainty. This value can also be used as the detection limit of the instrument.

The total uncertainty attributed is the combination of this random uncertainty and systematic uncertainty. Often a correction can be applied to compensate for the systematic effects. It is assumed that correction is done such that, after correction, the expected value of the error arising from a systematic effect is zero. A systematic uncertainty remains, however, characterized by the uncertainty in the correction.

There are many reasons why a measurement[1] is uncertain. For example, error components in satellite remote sensing may include terms such as

- instrument noise,
- error arising from simplifications in radiative transfer,
- calibration error,
- geolocation/interpolation error,
- error arising from the uncertainty in parameters used to derive the measurement.

An **uncertainty budget** is a list of random and systematic errors with estimates of the uncertainty they contribute to the measurement (preferably with information about how component uncertainties combine). Standard methods of error propagation (e.g. Hughes and Hase, 2010) are used to transform uncertainties into measurement units. The total uncertainty is the total combined accounting for any correlation between component errors.

In some cases the measurement process returns a vector of measurands. The error between the components of the measurand may not be independent so is represented by an uncertainty covariance matrix defined by

$$\text{uncertainty covariance matrix} = \begin{bmatrix} \langle \epsilon_1 \epsilon_1 \rangle & \langle \epsilon_1 \epsilon_2 \rangle & \cdots & \langle \epsilon_1 \epsilon_n \rangle \\ \langle \epsilon_2 \epsilon_1 \rangle & \langle \epsilon_2 \epsilon_2 \rangle & \cdots & \langle \epsilon_2 \epsilon_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle \epsilon_n \epsilon_1 \rangle & \langle \epsilon_n \epsilon_2 \rangle & \cdots & \langle \epsilon_n \epsilon_n \rangle \end{bmatrix}$$

[1] Measurement here is used to include satellite retrievals (estimates by some process of inversion) of measurands, although by some strict usage of "measurement", it is typically radiance that a sensor on a satellite actually measures.

where ϵ_i denotes the error on the ith measurand and <ab> denotes the expectation value of ab. If the measurands are independent then the off-diagonal terms are zero and the uncertainty on each measurand is given by the square-root of the corresponding diagonal element. For vector measurements, the uncertainty budget is a list of random and systematic errors with estimates of their associated uncertainty covariance matrices.

Two qualitative terms not defined in BIPM (2008) but commonly used to describe a measurement (e.g. Beers, 1957, Hughes and Hase, 2010) are precision and accuracy defined here as:

precision: a measurement which has a small random uncertainty is said to have high precision,

accuracy: a measurement which has a small systematic uncertainty is said to have high accuracy.

Validation of Measurements

Validation is the assessment of a measurement and the uncertainty attributed to it. This is principally achieved by **external validation,** i.e. comparison of a measurement to an independent measurement and assessment of their consistency relative to their estimated uncertainties. This independent estimate of the measurand is termed the **validation value**. The discrepancy is then defined as

discrepancy: the difference between the measurement and the validation value.

A small average discrepancy with respect to the root-sum-square of the measurement and validation value uncertainties is indicative of an accurate measurement, but could also result from a fortuitous cancellation of error terms.

For a small number of measurements it is possible to report individual discrepancies. However, for the large number of measurements typical of satellite remote sensing validation involves statistically characterising the discrepancies. There are often regimes of instrument behaviour for which uncertainties can be expected to differ, so it is usual to characterize discrepancies for the minimum number of regimes of consistent instrument behaviour. The choice of regimes could come from a cluster analysis of discrepancy (if the difference in regimes causes differences in systematic error), but more commonly comes from knowledge of the measurement process.

The statistical characterization of the discrepancies within a regime is made through three **quality parameters.** Consider the set of *n* measurements $\{x_1 \pm \square x_1, x_2 \pm \square x_2, x_3 \pm \square x_3, \ldots x_n \pm \square x_n\}$ of some quantity together with the set of validation values $\{v_1 \pm \square v_1, v_2 \pm \square v_2, v_3 \pm \square v_3, \ldots v_n \pm \square v_n\}$ made of the same quantity. The quality parameters are then:

Bias: the mean value of the discrepancy, i.e.

$$\text{bias} = b = \frac{\sum_{i=1}^{n}(x_i - v_i)}{n}$$

Chi-squared: the goodness of fit between the actual and estimated uncertainties of measurement and validation values, defined by

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - v_i)^2}{\delta x_i^2 + \delta v_i^2}$$

Stability: the change in bias with time defined as

$$\text{stability} = \frac{b(t+\Delta t)-b(t)}{\Delta t}$$

The expectation value of the bias is the sum of the residual systematic errors in the measurement and the validation value. The bias can only be attributed to the measurement if the residual systematic error in the validation value is known a priori. In an ideal case the bias would be zero.

The expected value for x^2 is unity. A value lower than this indicates the uncertainties attributed to the measurements or the validation values or both are too high. A value greater than unity indicates the uncertainties attributed to the measurements or the validation values or both are too low.

In the ideal case the stability would be zero over any timescale. In remote sensing the stability can display periodicity related to factors such as instrument drift or solar illumination of the satellite - both over an orbit and seasonally. It is suggested that the stability is estimated at the same temporal scale that any trends in the data are calculated.

It may be that the quality parameters are independent of the measurement magnitude and conditions of measurement and apply at all locations and times. In that case the three quality values adequately characterize the quality of measurement. More commonly, the quality values vary so a **validation table** is used to summarise the bias, x^2 and stability for regimes of consistent instrument behaviour.

In some case **internal validation** can be used to check reported uncertainty. Consider the situation where an instrument measures the same quantity under conditions where the reported uncertainty does not vary. Then the variability of the measurements should agree with the reported random uncertainty.

Comparing Measurements with a Model

Further understanding can be achieved through comparison of measurements with model output. In this approach, a model is sampled to give model values at the same place and time as the measurement values. The same three quality parameters can be calculated. However these caveats apply:

- the model error may not be reported and may have to be assumed,
- the bias cannot be attributed to the model or measurements without reference to additional information

An estimate of interpolation uncertainty must be included if the model reports results at different times and location from the measurements so that the model results are interpolated to the measurement location.

If the model is at a coarser resolution than the measurements an approach could be to compare the model value with a (weighted) average of the measurements. The fact that the systematic uncertainty is correlated needs to be accounted for if this approach is taken.

The statistical comparison of model and measurement data must account for bias due to sampling. For example a monthly time series comparison between model output and averaged measurements may show bias due to conditions, such as cloud coverage, under which measurements are not possible.

Guideline #2 - Common Table of Contents for Uncertainty Characterization Document Deliverable

The drafting team proposes a common table of contents for the "Uncertainty Characterization" deliverable:

1. Introduction

2. Definition of terms

-Error, Uncertainty, Uncertainty information, Uncertainty characterization, Validation, Accuracy, Precision, Stability, Representivity, Error co-variance matrix, etc.

3. Sources of errors

-Description of the sources of error contributing to uncertainty in the data products: qualitative-quantitative uncertainties, symmetric vs. asymmetric uncertainties, global vs. regionally differing uncertainties, error correlations, data pre-screening and other factors affecting the representativity of the data product, ...

4. Methodology to determine uncertainties

-Steps in algorithms, error propagation, analytical and empirical approaches to determining product uncertainties, ...

5. Documentation of uncertainties in the products

-Error budget analysis and results.

-e.g. uncertainty per land cover class, overall statistics on the uncertainties per product pixel, etc.

6. Guidelines for using the products

-how to use the data without introducing new uncertainties (e.g. level 2 to level 3 transition, data product representativity)

-how to use the uncertainty information

7. Conclusion

8. Bibliography

-e.g. peer reviewed publications on the methodology for characterising the uncertainties, cross reference to the validation reports, etc

10. Glossar

AATSR	Advanced Along Track Scanning Radiometer
AI	Aerosol Index
AMSRE	Advanced Microwave Scanning Radiometer - Earth Observing System
ATBD	Algorithm Theoretical Basis Document
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BADC	British Atmospheric Data Centre
BRDF	Bidirectional Reflectance Distribution Function
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CC4CL	Community Code 4 CLimate
COD	Cloud Optical Depth
CRE	Cloud Effective Radius
CTH	Cloud Top Height
CTP	Cloud Top Pressure
DISORT	Discrete Ordinates Radiative Transfer Program for a Multi-Layered Plane-Parallel Medium
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium Range Weather Forecast
EO	Earth Observation
FOV	Field Of View
GAC	Global Area Coverage – globally available AVHRR dataset with reduced resolution (4 km).
IR	Infrared
IWP	Ice Water Path
K	Kelvin

LWP	Liquid Water Path
MetOp	Meteorological Operational Satellite
MODIS	Moderate Resolution Imaging Spectroradiometer
NIR	Near Infrared
OE	Optimal Estimation
ORAC	Oxford RAL Aerosol and Cloud
PPSH	Plane Parallel Single layer Homogeneous cloud
RAL	Rutherford Appleton Laboratory
RMSE	Root Mean Square Error
RTM	Radiative Transfer Model
RTTOV	Radiative Transfer for TOVS
SST	Sea Surface temperature
VIS	Visible

11.References

Beers, Y., Introduction to the theory of error, Addison-Wesley, Massachusetts, 1957.

BIPM, Guide to the Expression of Uncertainty in Measurement (GUM), Bureau International des Poids et Mesures, 2008. (http://www.bipm.org/en/publications/guides/gum.html.)

Fischer, J. and M. Kollewe, 1994: Study of cloud top height determination using a coarse spatial resolution imaging spectrometer.- Final Report, ESA-ESTEC study, RFQ/3- 7241/91/NL/ BI.

Hughes, I.G., and T.P.A. Hase, Measurements and their uncertainties, Oxford University Press, Oxford, 2010.

O'Brien, D. and R.M. Mitchell, 1992: Error estimates for Retrieval of Cloud-Top Pressure Using Absorption in the A Band of Oxygen. J. Appl. Meteor., 31, 1179-1192

Preusker, R., and R. Lindstrot, 2009: Remote Sensing of Cloud-Top Pressure Using Moderately Resolved Measurements within the Oxygen A Band—A Sensitivity Study. J. Appl. Meteor. Climatol., 48, 1562–1574.

PVIR, Product Validation and Intercomparison Report (PVIR) – ESA Cloud_cci, Issue 5, Revision: 1, Date of Issue: March 2018, Available at: http://www.esa-cloud-cci.org/?q=documentation