



ESA Cloud_cci

Algorithm Theoretical Baseline Document v6.2
Community Cloud retrieval for Climate (CC4CL)
(Applicable to Cloud_cci version 3.0 products)



Issue 6 Revision 2

18/10/2019

Deliverable No.: D-2.1
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
Simon.Pinnock@esa.int

Document authors: Caroline A. Poulsen¹, Gareth E. Thomas¹, Richard Siddans¹, Adam Povey²,
Greg McGarragh², Cornelia Schlundt³, Stefan Stapelberg³, Martin Stengel³
and Roy G. Grainger²

¹ Space Science and Technology Department, RAL, Chilton, Didcot, OX11 0QX, UK

² Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, OX1 3PU, UK

³ Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach, Germany



Contents

1	Introduction	4
2	Retrieval system description	4
2.1	Cloud detection	4
2.2	Cloud type and phase determination	5
2.3	Spectral band adjustment (SBA)	8
2.4	Cloud property retrieval algorithm	10
2.4.1	Cloud / Atmosphere / Surface Model	11
2.4.2	Reflectance and transmission operators	12
2.4.3	Surface reflectance operators	14
2.4.4	Visible and near-IR RTM	14
2.4.5	Thermal-IR RTM	17
2.4.6	Derivatives of the forward model	17
2.4.7	Single layer Measurement vector and covariance	17
2.4.8	State vector and a priori constraint	18
2.4.9	Cloud Emissivity and microphysical retrievals at night and twilight	19
2.4.10	Cloud Albedo	19
2.4.11	Corrected CTT, CTP and CTH	20
2.4.12	Boundary layer inversion	21
2.4.13	Tropopause identification and temperature profile modification	21
2.4.14	ECMWF data	22
2.4.15	Quality Control	22
2.4.16	Modifications for ATSR-2/AATSR/SLSTR	22
2.4.17	Aerosol-Cloud (and sunglint) discrimination	22
2.5	Retrieval Performance	23
2.6	Changes compared to precursor CC4CL version	23
3	Input and output data	24
A	Bayesian scene identification	26



Document Change Record

Document, Version	Date	Changes	Originator
ATBD Version 5.1	12/09/2017	Released precursor version of ATBD applicable for Cloudcci dataset version 2	Caroline A. Poulsen, Gareth E. Thomas, Richard Siddans, Adam Povey, Greg McGarragh, Cornelia Schlundt, Stefan Stapelberg, Martin Stengel, Roy G. Grainger
ATBD Version 6.0	27/09/2018	Revision to make this document representing CC4CL used for processing of AVHRR-AM/PMv3 and ATSR2-AATSRv3 datasets. e.g. - renaming subsection on night retrievals to new subsection 2.4.9 including cloud emissivity retrieval Version submitted to ESA for review.	Caroline A. Poulsen, Gareth E. Thomas, Richard Siddans, Adam Povey, Greg McGarragh, Cornelia Schlundt, Stefan Stapelberg, Martin Stengel, Roy G. Grainger
ATBD Version 6.1	29/01/2019	Updates after ESA review of document version 6.0. E.g. - updating introduction section - updating section 2.1 on cloud detection ANN - updating section 2.2 on cloud phase ANN - introducing new subsection on spectral band adjustment - minor updates of subsection 2.4. - introducing new subsection 2.4.16 - modifying subsection 2.5 - introducing new subsection 2.6 on changes from previous to current CC4CL version.	Caroline A. Poulsen, Martin Stengel
ATBD Version 6.2	11/03/2019	Updates after ESA review of document version 6.1. E.g. - Introducing new subsection 2.4.17 on new Aerosol-Cloud (and sunglint) discrimination tests - Adding bullet point in Section 2.6 mentioning the usage of USGS 0.5km MODIS based Land Cover Type 2 data - Editing some text in Section 3 including mentioning options on using different snow/ice masks and on using higher resolved land/sea masks	Martin Stengel



1 Introduction

This document describes an optimal estimation retrieval scheme for the derivation of the properties of clouds from top-of-atmosphere (TOA) radiances measured by satellite-borne visible-IR radiometers as described in greater detail by Mcgarragh et al. [15], Sus el al. [31], Poulsen et al. [20] and Watts et al. [36]. The algorithm makes up part of the Community Code for CLimate (CC4CL) retrieval scheme (the other part, known as ORAC, performs aerosol retrievals and is described in Thomas et al. [34]). CC4CL was originally designed for application to SEVIRI and was subsequently adapted for the ATSR instrument, which was used to produce the GRAPE (Global Retrieval of ATSR cloud Parameters and Evaluation) data set. Within the Cloud_cci framework, the algorithm has been applied to AVHRR and MODIS using the ‘heritage’ set of channels (i.e. channels equivalent to the 0.67, 0.87, 1.6, 3.7, 11 and 12 μm channels of AVHRR) to produce a long-term climate data record with global coverage. The algorithm has also been adapted to retrieve cloud properties from Himawari, GOES and SEVIRI instruments. The algorithm can ingest most visible and Infrared wavelengths and considerable performance enhancement can be achieved when additional channels are used in the retrieval for example improved cloud top height is achieved when the CO₂ slicing channels are used. In the text that follows, the description will often refer to AATSR but is equally applicable to instruments with similar channel definitions.

Specific features of this algorithm include:

- A full implementation of the optimal estimation framework described by Rodgers [22], enabling rigorous error propagation and inclusion of *a priori* knowledge.
- A common retrieval algorithm over both land and ocean, with only the *a priori* constraint on the surface reflectance differing between the two.
- Consistent and simultaneous retrieval of all cloud parameters in the visible and infrared.

2 Retrieval system description

CC4CL consists of three main components: (1) cloud detection (2) cloud typing and (3) the retrieval of cloud properties based on OE technique. These components are described in more detail in the following sections.

2.1 Cloud detection

The cloud mask is based on artificial neural networks (ANNs). Three ANNs were trained: for day (using visible (VIS), near-infrared (NIR) and infrared (IR) information), for night/twilight (using NIR and IR channels), and again for night/twilight with different channel input (IR channels only). The ANNs have been trained based on collocations between AVHRR NOAA-19 measurements and CALIPSO COD data. As a result a pseudo CALIPSO optical depth (ANNCOD) is computed (mimicked) and converted into a 2 bit cloud mask by applying thresholds.

The used ANN’s are multilayer perceptrons consisting of 1 input layer, one hidden layer with about 50 neurons and one output layer. The training and test data sets are based on collocated CALIPSO COD data (CAL_LID_L2_05kmClay-Prov-V3-01, total optical thickness at 532nm) and Cloud_cci AVHRR GAC Level-1C measurements. The Day ANN was trained with VIS and IR data. Compared to earlier versions, Ch3b (3.7 μm) of AVHRR was now included Day ANN. For NOAA-17, METOP-A and partly NOAA-16 a separate ANN was trained including 1.6 μm instead of 3.7 μm during day. The 2 night ANN’s have been trained both with and without the 3.7 μm NIR channel information. This avoids the misclassification of very cold clouds and/or surfaces because of very low 3.7 μm channel signal-to-noise-ratios. The channel input is again summarizes in table 1.



Overall, 5 years of globally and seasonally representative collocation data between NOAA-19 and CALIPSO have been chosen as training dataset. Prior to that, all collocated CALIPSO COD's greater 1 have been set to 1. In a final step, a simple viewing-angle correction is applied to the retrieved ANNCOD which tends to increase with increasing viewing-angle. For the generation of the final 2-bit cloud mask, suitable thresholds have been found and applied to the ANNCOD.

Cloud detection thresholds:

Table 2 lists the threshold applied to determine the binary cloud decision from ANNCOD. The thresholds depend on illumination and surface type.

As the ANNs were trained using NOAA-19 AVHRR measurements, a radiances adjustment has been determined for AATSR and MODIS to make them mimic AVHRR. The adjustment includes slopes and offsets based on synthetic measurements with details given in Section 2.3.

Cloud detection uncertainty:

Based on the training set. The binary results (after applying ANN and thresholds) were compared against the CALIOP cloud mask and detection scores calculated. The scores show a clear dependence on the distance between the ANNCOD and the respective threshold, which allows a approximation of the cloud detection uncertainty on pixel level based on these two values. Figure 1 shows the probability of incorrect cloud mask as a function of normalized ANNCOD-threshold distance.

Table 1: Measurement input to the trained artificial neural network for cloud detection (ANN_{mask}), used for different illumination conditions: daytime, twilight and night-time. The subscript in the table's headline corresponds to the approximate central wavelengths of the channels: $0.6\mu\text{m}$, $0.9\mu\text{m}$, $1.6\mu\text{m}$, $3.7\mu\text{m}$, $10.8\mu\text{m}$, $12.0\mu\text{m}$. In addition to the measurement input, all ANNs require surface temperature, a snow-ice flag and a land-sea flag as input. R=reflectance, BT=brightness temperature

ANN_{mask}	R _{0.6}	R _{0.9}	R _{1.6}	R _{3.7}	BT _{3.7}	BT _{10.8}	BT _{12.0}	BT _{10.8} -BT _{12.0}	BT _{10.8} -BT _{3.7}
Day _{3.7}	✓	✓	-	✓	-	✓	✓	✓	-
Day _{1.6}	✓	✓	✓	-	-	✓	✓	✓	-
Day _{backup}	✓	✓	-	-	-	✓	✓	✓	-
Twilight _{3.7}	-	-	-	-	✓	✓	✓	✓	✓
Twilight _{backup}	-	-	-	-	-	✓	✓	✓	-
Night _{3.7}	-	-	-	-	✓	✓	✓	✓	✓
Night _{backup}	-	-	-	-	-	✓	✓	✓	-

2.2 Cloud type and phase determination

A new cloud typing algorithm has been implemented in the preprocessing to select a cloud phase based on the cloud type it selects:

Cloud type:

- The following categories indicate the liquid cloud phase:
 - fog,
 - warm liquid water clouds, and
 - supercooled-mixed-phased clouds.



Table 2: Thresholds used to convert the output of the cloud mask ANNs to a binary cloud mask. Thresholds depend on illumination conditions and surface types.

Illumination	Surface type	Threshold
Day	Sea ice	0.4
Day	Land ice	0.3
Day	Sea	0.25
Day	Land	0.3
Night	Sea ice	0.45
Night	Land ice	0.35
Night	Sea	0.25
Night	Land	0.3
Twilight	Sea ice	0.5
Twilight	Land ice	0.35
Twilight	Sea	0.35
Twilight	Land	0.45

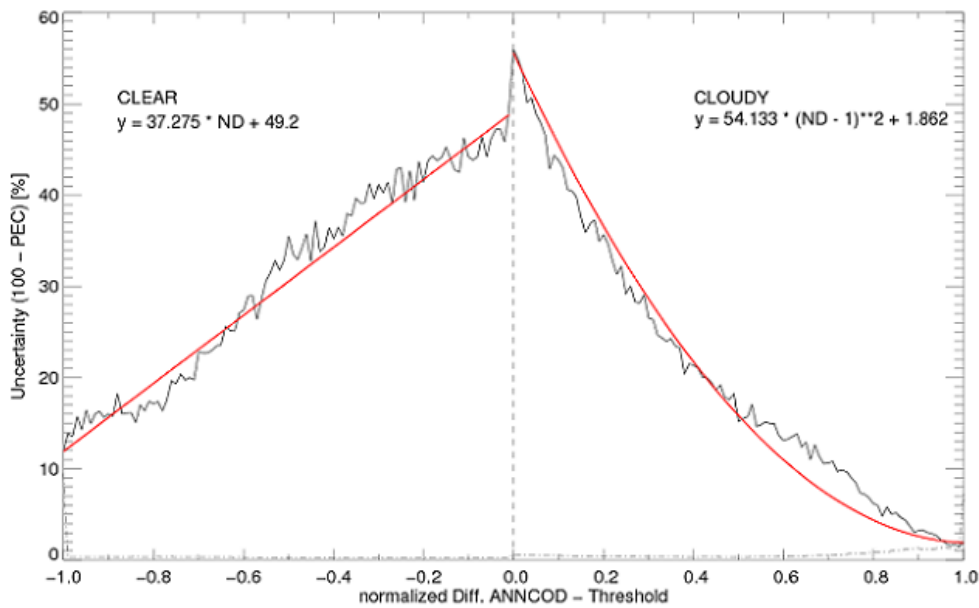


Figure 1: Neural network cloud mask uncertainty. Figure taken from Sus et al. [31]

- The following categories indicate the ice cloud phase:
 - opaque ice clouds/deep convection,
 - non-opaque high ice clouds (e.g. cirrus),
 - cloud overlap¹ (e.g. multiple cloud layers), and
 - probably-opaque ice clouds (e.g. neither 1.6 μm nor 3.7 μm channel is available during night for AVHRR; 12 μm channel is missing during night for AATSR).

¹Assuming that a scene with cloud overlap consists of a semitransparent ice cloud that overlaps a cloud composed of liquid water droplets [18].



Usually, CC4CL performs separate retrievals for each phase and selects the appropriate phase as that with the lowest retrieval cost (eqn. 1). If desired, CC4CL can now process only once assuming the phase identified above.

A detailed description of daytime spectral tests using AVHRR, MODIS and VIIRS observations is given in Pavolonis and Heidinger [18] and Pavolonis et al. [19] along with inter-comparison results. The multispectral algorithm used in CC4CL is based on the threshold approach described in these papers but can retrieve in the absence of visible/near-IR measurements (e.g. at night) and includes a new threshold test required for AVHRR and A(A)TSR measurements (i.e. the probably-opaque ice clouds).

To briefly summarize the main features of the algorithm, the cloud type retrieval requires as input satellite imager data at 0.65, 1.6(3a)/3.75(3b), 10.8 and 12 μm . If both 1.6 and 3.75 μm data are available, the 3.75 μm reflectances are used.² Pavolonis et al. [19] have analyzed cloud typing results based on channel 3a or 3b and concluded that both algorithms produce nearly identical results except for certain thin clouds and at cloud edges. More often than the 3b algorithm, the retrieval using 3a tends to misclassify the thin edges of some low and mid-level clouds as cirrus and opaque ice.

The algorithm also requires a cloud mask as only satellite pixels which are fully covered by cloud should be processed. Pavolonis et al. [19] have shown that if any clear pixels are passed to the cloud typing code, they would mostly be identified as warm liquid water or supercooled-mixed-phase cloud depending on the brightness temperature at 11 μm over most surface types and as opaque ice cloud over cold snow/ice covered surfaces.

Finally, the surface type of each single satellite pixel is required because different threshold values are applied depending on the underlying surface, particularly for water, desert and snow/ice.

Cloud phase:

While in the previous CC4CL version, the cloud types were just converted to a binary cloud phase, in the present CC4CL version an ANN for cloud phase was developed and implemented.

Three ANNs were trained: for day (using visible (VIS), near-infrared (NIR) and infrared (IR) information), for night/twilight (using NIR and IR channels), and again for night/twilight with different channel input (IR channels only). The ANNs have been trained based on collocations between AVHRR NOAA-19 measurements and CALIPSO CPH (cloud phase) data. The output of the phase ANN is a pseudo CALIPSO cloud phase although giving continuous values between 0 (liquid) and 1 (ice) which requires the application of thresholds posterior to infer 2 bit cloud phase information mask by applying thresholds.

As for cloud detection, the used ANN's are multilayer perceptrons consisting of 1 input layer, one hidden layer with about 50 neurons and one output layer. The training and test data sets are based on 5 years of collocated CALIPSO CPH data (CAL_LID_L2.05kmCLay-Prov-V3-01) and NOAA-19 AVHRR GAC Level-1C measurements. The Day ANN was trained with VIS, NIR and IR data. As defaults the 3.7 μm channel is expected; for NOAA-17, METOP-A and partly NOAA-16 a separate ANN was trained including 1.6 μm instead of 3.7 μm during day. The 2 night ANN's have been trained both with and without the 3.7 μm NIR channel information. The channel input is again summarized in table ???. A simple viewing-angle correction is applied to the ANN output. For the generation of the final 2-bit cloud phase, suitable thresholds have been found and applied.

Cloud phase thresholds:

Table 4 lists the threshold applied to determine the binary cloud phase from ANN output. The thresholds depend on illumination and surface type.

As the ANNs were trained using NOAA-19 AVHRR measurements, a radiances adjustment has been de-

²AVHRR sensors mounted on NOAA-7, 9, 11, 12 and 14 have only five channels (3a is missing). NOAA-15 through NOAA-19 as well as MetOp-A and MetOp-B have six channels, but only 5 channels are transmitted simultaneously, such that there is a potential switch in availability between 3a and 3b for certain time periods and sensors (generally 3a during the day and 3b during the night).



terminated for AATSR and MODIS to make them mimic AVHRR. The adjustment includes slopes and offsets based on synthetic measurements with details given in Section 2.3.

Cloud phase uncertainty:

Cloud phase uncertainty is done in an identical fashion as for cloud detection (see above).

Table 3: Measurement input to the trained artificial neural network for cloud phase determination (ANN_{phase} , used for different illumination conditions: daytime, twilight and night-time. The subscript in the table's headline corresponds to the approximate central wavelengths of the channels: $0.6\mu\text{m}$, $0.9\mu\text{m}$, $1.6\mu\text{m}$, $3.7\mu\text{m}$, $10.8\mu\text{m}$, $12.0\mu\text{m}$. In addition to the measurement input, all ANNs require a surface type flag containing the values 0:sea,1:land,2:desert,3:sea-ice,4:snow.

ANN_{phase}	$R_{0.6}$	$R_{0.9}$	$R_{1.6}$	$R_{3.7}$	$BT_{3.7}$	$BT_{10.8}$	$BT_{12.0}$	$BT_{10.8}-BT_{12.0}$	$BT_{10.8}-BT_{3.7}$
Day _{3.7}	✓	✓	-	✓	-	✓	✓	✓	-
Day _{1.6}	✓	✓	✓	-	-	✓	✓	✓	-
Twilight	-	-	-	-	✓	✓	✓	✓	✓
Night	-	-	-	-	✓	✓	✓	✓	✓

Table 4: Thresholds used to convert the output of the cloud phase ANNs to a binary cloud phase. Thresholds depend on illumination conditions and surface types.

Illumination	Surface type	Threshold
Day	Sea ice	0.5
Day	Land ice	0.7
Day	Sea	0.55
Day	Land	0.7
Night	Sea ice	0.7
Night	Land ice	0.6
Night	Sea	0.5
Night	Land	0.65
Twilight	Sea ice	0.7
Twilight	Land ice	0.9
Twilight	Sea	0.65
Twilight	Land	0.50

2.3 Spectral band adjustment (SBA)

To mimic NOAA-19 AVHRR, for which the cloud detection and cloud phase determination were developed and fine-tuned spectral band adjustment factors (SBA) were determined. The SBAs (slope and offset) were inferred from linearly fitting synthetic measurements between pairs of sensors (sensor X and NOAA19 AVHRR). The synthetic measurements were calculated from a set of SCIAMACHY and IASI orbits. SCIAMACHY and IASI provide hyperspectral measurements throughout the visible (SCIAMACHY) and infrared (IASI) part of the electromagnetic spectrum. Convolving the spectral response functions (SRF) of AVHRR channels $0.6\mu\text{m}$, $0.9\mu\text{m}$, $10.8\mu\text{m}$ and $12.0\mu\text{m}$ over the SCIAMACHY and IASI spectra, synthetic AVHRR measurements in each SCIAMACHY and IASI footprint are inferred building a set of points which is the basis for the linear fitting. This SBAs are applied prior to cloud detection and cloud typing procedures. Near-infrared channels at $1.6\mu\text{m}$ and $3.7\mu\text{m}$ are not adjusted as the SCIAMACHY and IASI spectra do not cover the full AVHRR



SRF of these channels. All inferred SBAs are given in Tables 5 and 6. In contrast to cloud detection and phase determination, the OE retrieval uses the individual SRFs thus SBAs are not required for the OE retrieval.

Table 5: Linear regression coefficients (slope and offset) applied as spectral band adjustment to either measured reflectances (R) or brightness temperature of all AVHRR-heritage channels (Ch) and sensors (used so far) to mimic NOAA-19 AVHRR. The subscript in the table's headline corresponds to the approximate central wavelengths of the channels: 0.6 μ m, 0.9 μ m, 1.6 μ m, 3.7 μ m, 10.8 μ m, 12.0 μ m. Reflectances in channels 0.6 μ m, 0.9 μ m and 1.6 μ m are generally not used in twilight and night conditions. Table adopted from Stengel et al. [30]

	R _{0.6} slope / offset	R _{0.9} slope / offset	R _{1.6} slope / offset	BT _{3.7} slope / offset	BT _{10.8} slope / offset	BT _{12.0} slope / offset
day						
noaa-5	1.009 / -0.174	1.011 / -0.102	1.000 / 0.000	0.985 / 3.681	0.999 / 0.236	1.000 / 0.000
noaa-6	1.012 / -0.027	1.003 / -0.052	1.000 / 0.000	0.976 / 6.078	0.999 / 0.300	1.000 / 0.000
noaa-7	1.009 / -0.036	1.007 / -0.007	1.000 / 0.000	1.010 / -2.531	1.000 / -0.198	0.991 / 1.991
noaa-8	1.010 / -0.009	1.003 / -0.048	1.000 / 0.000	0.985 / 3.277	0.999 / 0.201	1.000 / 0.000
noaa-9	1.009 / -0.013	1.006 / 0.011	1.000 / 0.000	1.014 / -3.535	1.000 / -0.215	0.988 / 2.770
noaa-10	1.010 / -0.044	1.005 / -0.020	1.000 / 0.000	1.001 / -0.462	0.999 / 0.469	1.000 / 0.000
noaa-11	1.009 / -0.010	1.005 / -0.012	1.000 / 0.000	1.008 / -2.286	1.000 / -0.170	0.989 / 2.443
noaa-12	1.008 / 0.012	1.005 / -0.029	1.000 / 0.000	0.986 / 3.190	1.001 / -0.212	0.994 / 1.383
noaa-13	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000
noaa-14	1.008 / 0.016	1.011 / -0.026	1.000 / 0.000	0.992 / 2.122	1.001 / -0.446	0.995 / 1.081
noaa-15	1.005 / -0.027	1.006 / 0.047	1.000 / 0.000	1.012 / -3.984	1.000 / -0.136	0.990 / 2.145
noaa-16	1.006 / -0.039	1.009 / 0.057	1.000 / 0.000	1.015 / -4.572	1.000 / -0.095	0.997 / 0.561
noaa-17	1.004 / -0.021	1.008 / 0.053	1.000 / 0.000	0.998 / 0.037	1.000 / -0.185	0.990 / 2.112
noaa-18	1.002 / -0.013	1.015 / 0.066	1.000 / 0.000	0.994 / 1.422	1.000 / -0.214	0.997 / 0.626
noaa-19	1.000 / 0.000	1.000 / -0.000	1.000 / 0.000	1.000 / 0.000	1.000 / -0.000	1.000 / 0.000
noaa-20	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000
metop-1	1.006 / -0.033	1.011 / 0.033	1.000 / 0.000	0.998 / 0.899	1.000 / -0.165	0.990 / 2.300
metop-2	1.006 / -0.033	1.006 / 0.038	1.000 / 0.000	1.013 / -3.203	1.000 / -0.196	0.991 / 1.981
modis-t	0.978 / 0.202	0.897 / -0.004	1.000 / 0.000	0.975 / 5.113	0.998 / 0.571	0.985 / 3.596
modis-a	0.979 / 0.201	0.897 / -0.005	1.000 / 0.000	0.976 / 4.413	0.998 / 0.674	0.985 / 3.614
atsr2	0.959 / 0.259	0.894 / -0.012	1.000 / 0.000	1.045 / -9.727	0.999 / 0.435	1.000 / 0.034
aatsr	0.956 / 0.259	0.893 / -0.013	1.000 / 0.000	1.005 / -1.764	0.999 / 0.268	1.003 / -0.753



Table 6: As table 5 but for twilight and night conditions.

	R _{0.6} slope / offset	R _{0.9} slope / offset	R _{1.6} slope / offset	BT _{3.7} slope / offset	BT _{10.8} slope / offset	BT _{12.0} slope / offset
twilight & night						
noaa-5	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.991 / 1.398	0.999 / 0.345	1.000 / 0.000
noaa-6	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.985 / 2.452	0.999 / 0.394	1.000 / 0.000
noaa-7	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.008 / -1.902	1.000 / -0.194	0.992 / 1.786
noaa-8	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.989 / 1.610	0.999 / 0.292	1.000 / 0.000
noaa-9	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.011 / -2.387	1.000 / -0.243	0.989 / 2.500
noaa-10	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / -0.424	0.998 / 0.565	1.000 / 0.000
noaa-11	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.006 / -1.500	1.000 / -0.178	0.990 / 2.184
noaa-12	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.991 / 1.194	1.000 / -0.111	0.994 / 1.218
noaa-13	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000
noaa-14	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.996 / 0.290	1.001 / -0.427	0.996 / 0.945
noaa-15	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.006 / -1.306	1.000 / -0.109	0.991 / 1.903
noaa-16	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.008 / -1.578	1.000 / 0.022	0.997 / 0.511
noaa-17	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.999 / -0.071	1.000 / -0.173	0.992 / 1.877
noaa-18	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.996 / 0.366	1.000 / -0.209	0.997 / 0.542
noaa-19	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000
noaa-20	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000
metop-1	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.000 / -0.040	1.001 / -0.218	0.991 / 2.053
metop-2	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.011 / -2.633	1.000 / -0.170	0.992 / 1.744
modis-t	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.979 / 4.234	0.997 / 0.782	0.987 / 2.963
modis-a	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	0.979 / 4.277	0.997 / 0.876	0.987 / 2.978
atsr	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.038 / -7.470	0.999 / 0.521	1.000 / 0.018
aatsr	1.000 / 0.000	1.000 / 0.000	1.000 / 0.000	1.002 / -0.548	0.999 / 0.265	1.003 / -0.667

2.4 Cloud property retrieval algorithm

Although the algorithm described here is applicable to measurements in a wide range of visible to thermal infrared atmospheric window channels, within the Cloud_cci project it is being applied to the so-called *heritage channels* first used on the AVHRR series of instruments. These are six channels centred around 0.67, 0.87, 1.6, 3.7, 11, and 12 μm (although there can be considerable differences in the widths and central wavelengths of filter band-passes in different instruments). These channels are sensitive in different ways to the macro and microphysical properties of cloud. For example, the infrared channels compliment the visible channels in the case of optically thin clouds. However, the observations are not sensitive to every aspect of the three-dimensional distribution of all relevant cloud properties and no single channel is uniquely sensitive to a specific cloud property.

We approach the problem of extracting useful information on cloud as an inverse problem. A forward model (FM) is defined which applies a radiative transfer model (RTM) to simulate satellite radiances based on a parameterized cloud / atmosphere / surface model (CM) and the prescribed observing conditions. An inverse or retrieval model (RM) is then used to obtain the cloud parameters which give the best fit between the model predicted and observed radiances, taking into account measurement uncertainties and relevant prior knowledge. This inverse problem is solved using the optimal estimation method [22] (OEM).

The basic principle of the OEM is to maximise the probability of the retrieved state, conditional on the value of the measurements and any *a priori* knowledge. Formally, it maximises the conditional probability $P = P(\vec{x} | \vec{y}, \vec{x}_a)$ with respect to the values of the measurement vector \vec{y} , state vector \vec{x} , and *a priori* estimate



of the state \vec{x}_a (i.e. the most likely state prior to considering the measurements). It is assumed that errors in the measurements, forward model and *a priori* parameters are normally distributed with zero mean and covariances given by \mathbf{S}_y and \mathbf{S}_a , respectively. The solution state is found by minimising the cost function J :

$$J(\vec{x}) = [\vec{y}(\vec{x}) - \vec{y}_m] \mathbf{S}_y^{-1} [\vec{y}(\vec{x}) - \vec{y}_m]^T + (\vec{x} - \vec{x}_a) \mathbf{S}_a^{-1} (\vec{x} - \vec{x}_a)^T. \quad (1)$$

Starting from some initial guess of the state and linearising the forward model, the gradient of the cost function is estimated. Using that, a state is selected which is predicted to have lower cost. The Levenberg-Marquart [14, 11] scheme is used to perform the minimisation. The procedure is iterated until the change in cost between iterations is less than $0.05m$, where m is the length of \vec{y} , (called convergence) or the retrieval is abandoned after 40 iterations.

If the *a priori* and measurement uncertainties are well represented by their respective covariances, the value of the cost function at solution is expected to be sampled from a χ^2 distribution with degrees of freedom (approximately) equal to the total number of elements in the measurement and state vectors. Hence, J at convergence provides a measure of the likelihood of the solution-state being consistent with observations and prior knowledge.

For retrievals which satisfactorily converge, i.e. converge to a minimum cost which is consistent with measurement and prior uncertainties, the uncertainty on the estimated state parameters is described by the solution covariance

$$\mathbf{S}_x = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1}, \quad (2)$$

where \mathbf{K} contains the derivatives of the forward model with respect to each solution state parameter:

$$K_{i,j} = \frac{\partial y_i}{\partial x_j}. \quad (3)$$

2.4.1 Cloud / Atmosphere / Surface Model

The retrieval forward model can be considered to consist of three components: a scattering cloud layer is located within a clear-sky atmosphere over a surface of known reflectance/emissivity. The clear-sky atmosphere is defined by temperature and humidity profiles taken from ECMWF ERA Interim analyses [7]. As only window channels are being used in the retrieval, the influence of variations in trace gas concentrations, as well as the uncertainties in ECMWF water vapour profiles, are well within the measurement noise.

The surface is characterised by a bidirectional reflectance distribution function (BRDF) which is computed differently for ocean and land surface. The BRDF over ocean is computed using the methodology outlined by [25] which includes 3 components:

$$\rho(\theta_0, \theta_v, \phi, \lambda, u, v) = \rho_{sg}(\theta_0, \theta_v, \phi, \lambda, u, v) + \rho_{wc}(\lambda, u, v) + \rho_{ul}(\theta_0, \theta_v, \lambda, C), \quad (4)$$

where ρ_{sg} is the sun-glint off wave facets [5], ρ_{wc} is the reflectance from surface foam, so-called ‘‘whitecaps’’ [9], and ρ_{ul} is the scattering from the within the water, so-called ‘‘underlight’’ [16]. In addition, physical parameters include the horizontal wind vector u and v (m/s) and the ocean pigment concentration C (mg/m³).

The BRDF over land is a weighted sum of an isotropic kernel (unity) and two BRDF kernels [35]

$$\rho(\theta_0, \theta_v, \phi, \lambda) = f_{iso}(\lambda) + f_{vol}(\lambda) K_{vol}(\theta_0, \theta_v, \phi) + f_{geo}(\lambda) K_{geo}(\theta_0, \theta_v, \phi), \quad (5)$$

where $K_{vol}(\theta_0, \theta_v, \phi)$ is known as the the Ross-thick kernel which parameterises *volumetric* scattering of leaves in dense vegetation and $K_{geo}(\theta_0, \theta_v, \phi)$ is the Li-sparse kernel which parameterises *geometric* shadowing in sparsely wooded vegetation. The weights $f(\lambda)$ are provided by the 0.05° MODIS MCD43C1 BRDF auxiliary input discussed in section 3.



For the infrared channels the surface is assumed to have an emissivity of unity over the ocean, while the CIMSS global land emissivity database is used [28]. The temperature of the surface is a retrieved parameter (see section 2.4.8).

Each measurement pixel is considered to be either fully cloudy or clear. The algorithm does provide the capability of retrieving the cloud-filled fraction of a pixel, but it has been found that the heritage channels do not provide sufficient information to distinguish thin but complete cloud cover from thick but partial cover. Cloud is assumed to be a single, plane-parallel layer of either liquid or ice particles. The layer is assumed to be (geometrically) infinitely thin and is placed within the clear-sky atmosphere model. The cloud layer is parametrised in terms of the following retrieved quantities:

- The cloud phase, i.e. ice or liquid.
- The effective radius r_{eff} of the cloud particle size distribution.
- The total (vertically integrated) optical depth τ of the cloud at a fixed wavelength of $0.55 \mu\text{m}$.
- The cloud top pressure p_c .

Size distributions for ice and liquid cloud are defined as a function of only r_{eff} , which defines the shape of the modelled size distribution, and τ , implicitly defining the total number of particles. For ice clouds, single-scattering properties (extinction coefficient, single-scattering albedo and phase function) are taken from Yang et al. [37]. Single-scattering properties of liquid cloud are derived by Mie theory assuming a modified gamma size distribution of particle radius r such that

$$n(r) = 2.373 r^6 \exp\left(\frac{-6r}{r_m}\right), \quad (6)$$

where r_m is the mode radius of the distribution. The radiatively significant effective radius r_{eff} is given by

$$r_{\text{eff}} = \frac{\int_0^\infty r \pi r^3 n(r) dr}{\int_0^\infty \pi r^2 n(r) dr}. \quad (7)$$

This approach reduces the complexity of cloud to a simple model with parameters which can be distinguished using the heritage channels. The visible channel radiances are predominantly controlled by the cloud optical depth. Near-IR channels are also sensitive to particle size and phase due to the dependence on size of the single-scattering albedo in that spectral range and the associated differences between ice and liquid phase particles. Thermal channels predominantly provide information on cloud top pressure (via the dependence of the cloud thermal emission on the atmospheric temperature profile). It should be appreciated that all channels are sensitive, to a greater or lesser extent, to all parameters (dependant on the scene).

This simple model cannot represent all aspects of cloud three-dimensional structure. In the ideal case, the retrieved parameters should correspond to vertical (over the profile) and horizontal (over the scene) averages of the “true” cloudy properties. However, there are classes of clouds, particularly those with strong vertical variations in particle size and phase, for which this model cannot predict radiances consistent with observations in all channels. Such conditions can be identified by checking that the retrieval converges with satisfactory cost.

2.4.2 Reflectance and transmission operators

The next step in the forward model is the prediction of transmission and reflectance operators for an atmosphere *without* molecular absorption. This calculation is based on solar and viewing geometry, the molecular scattering optical thickness τ_{ms} and single-scattering phase function $P_{\text{ms}}(\lambda, \Theta)$, the optical thickness $\tau_a(\lambda)$, the single-scattering albedo $\omega_a(\lambda)$ and the single-scattering phase function $P_a(\lambda, \Theta)$. For performance reasons this



calculation is look-up table (LUT) based from which the values for an arbitrary set of geometric and optical parameters may be interpolated. The vertices of the LUTs are computed with the DIScrete Ordinate Radiative Transfer (DISORT) software package [32]. It is important to note that this step, although slow, is performed off-line and the resulting look-up tables (LUTs) are static.

DISORT is a thoroughly documented and widely used general purpose algorithm for the calculation of time-independent radiative transfer. The DISORT algorithm solves the equation for the transfer of monochromatic light at wavelength λ in a medium with absorption and multiple scattering, including solar and thermal sources. The radiative transfer equation is written as

$$\mu \frac{dL_\lambda(\tau_\lambda, \mu, \phi)}{d\tau} = L_\lambda(\tau_\lambda, \mu, \phi) - J_\lambda(\tau_\lambda, \mu, \phi), \quad (8)$$

where $L_\lambda(\tau_\lambda, \mu, \phi)$ is the intensity along direction μ, ϕ (where μ is the cosine of the zenith angle and ϕ is the azimuth angle) at optical depth τ_λ measured perpendicular to the surface of the medium. $J_\lambda(\tau_\lambda, \mu, \phi)$ is the source function, which can include solar and thermal sources.

It should be noted that DISORT still makes some important approximations, which can limit its accuracy in certain circumstances. The most important of these are:

- It assumes a plane parallel atmosphere, which makes it inapplicable at viewing or zenith angles above approximately 75° , where the curvature of the Earth has a significant influence on radiative transfer.
- It is a one-dimensional model, so cannot reproduce the effects of horizontal gradients in the scattering medium. This is important where strong gradients exist, such as near cloud edges.
- It does not model polarisation effects and hence cannot be used to model measurements made by instruments which are sensitive to polarisation and does not take into account the polarisation introduced into the diffuse component of radiance by molecular scattering.

DISORT is provided with solar and instrument geometry, the molecular scattering and cloud radiative properties at the vertexes. The transmission and reflectance of the atmosphere is computed for both direct beam and diffuse radiation sources separately. The calculations are performed quasi-monochromatically, i.e. a single radiative transfer calculation is performed for each channel. It is the input optical properties that are convolved to the instrument's response function for a particular channel. These calculations produce six LUTs for each channel:

- $R_{bb}(\theta_0, \theta_v, \phi)$: the bidirectional reflectance of the cloud.
- $T_{bb}^\downarrow(\theta_0)$: the downward direct transmission of the cloud of the direct solar beam.
- $T_{bb}^\uparrow(\theta_v)$: the upward direct transmission of the cloud into the viewing direction.
- $T_{bd}^\downarrow(\theta_0)$: the downward diffuse transmission of the cloud, as illuminated by the direct solar beam.
- $T_{db}^\uparrow(\theta_v)$: the upward diffuse transmission of the cloud, as viewed from a specific direction.
- R_{dd} : the bi-hemispherical reflectance of the cloud.

Here, a \downarrow denotes transmission from the top to the bottom of the atmosphere, while \uparrow indicates the reverse. Dependence on the solar zenith, viewing zenith and relative azimuth angles are denoted by θ_0 , θ_v and ϕ respectively. The pairs of b and d subscripts denote the type of radiation each term operates on and produces; for example $T_{bd}^\downarrow(\lambda, \theta_0)$ operates on the direct beam (b) of solar radiation, and produces the diffuse radiation (d) that results at the bottom of the atmosphere. Each of these tables contains tabulated transmission or reflectance (depending on the table) values for each of the ten equally spaced solar and/or sensor zenith angles, eleven equally spaced relative azimuth angles ($R_{bb}(\lambda, \theta_0, \theta_v, \phi)$ only), eighteen $0.55 \mu\text{m}$ optical depths and twenty three effective radii.



2.4.3 Surface reflectance operators

The CC4CL forward model works on the assumption that the surface BRDF can be parameterized by four reflectance terms:

1. The bidirectional reflectance, $\rho_{bb}(\lambda, \theta_0, \theta_v, \phi)$. This is the reflectance of the surface to direct beam illumination at θ_0 , as viewed from a specific direction θ_v . It is the reflectance that would be observed by a satellite instrument in the absence of an atmosphere.
2. The directional-hemispheric reflectance $\rho_{bd}(\lambda, \theta_0)$. This is the fraction of incoming direct beam illumination at θ_0 that is reflected across all viewing angles. This is also referred to as the *black-sky albedo*.
3. The hemispheric-directional reflectance $\rho_{db}(\lambda, \theta_v)$. This is the reflectance of the surface to purely diffuse illumination, as viewed from a specific direction θ_v .
4. The bi-hemispheric reflectance $\rho_{dd}(\lambda)$. This is the reflectance of the surface to purely diffuse illumination, across all viewing directions. This is also referred to as the *white-sky albedo*.

The first term $\rho_{bb}(\theta_0, \theta_v, \phi)$ is computed directly from the BRDF. The three other terms are derived from the BRDF integrated over solar and/or viewing geometry written as

$$\begin{aligned}\rho_{bd}(\lambda, \theta_0) &= \frac{\int_0^{2\pi} \int_0^{\pi/2} \rho_{bb}(\lambda, \theta_0, \theta_v, \phi) \cos \theta_v \sin \theta_v d\theta_v d\phi}{\int_0^{2\pi} \int_0^{\pi/2} \cos \theta_v \sin \theta_v d\theta_v d\phi} \\ &= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \rho_{bb}(\lambda, \theta_0, \theta_v, \phi) \cos \theta_v \sin \theta_v d\theta_v d\phi,\end{aligned}\quad (9)$$

$$\begin{aligned}\rho_{db}(\lambda, \theta_v) &= \frac{\int_0^{2\pi} \int_0^{\pi/2} \rho_{bb}(\lambda, \theta_0, \theta_v, \phi) \cos \theta_v \sin \theta_v d\theta_0 d\phi}{\int_0^{2\pi} \int_0^{\pi/2} \cos \theta_v \sin \theta_v d\theta_0 d\phi} \\ &= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \rho_{bb}(\lambda, \theta_0, \theta_v, \phi) \cos \theta_v \sin \theta_v d\theta_0 d\phi,\end{aligned}\quad (10)$$

$$\begin{aligned}\rho_{dd}(\lambda) &= \frac{\int_0^{\pi/2} \rho_{bd}(\lambda, \theta_0) \cos \theta_0 \sin \theta_0 d\theta_0}{\int_0^{\pi/2} \cos \theta_0 \sin \theta_0 d\theta_0} \\ &= 2 \int_0^{\pi/2} \rho_{bd}(\lambda, \theta_0) \cos \theta_0 \sin \theta_0 d\theta_0.\end{aligned}\quad (11)$$

2.4.4 Visible and near-IR RTM

Each short wave channel measures the radiance in the instrument's field-of-view, defined by the solid angle Δ_{FOV} . Each channel also has a relative spectral response $\varrho(\lambda)$ within a wavelength interval $[\lambda_1, \lambda_2]$ and has zero response outside this band. Under these conditions the radiance measured by the instrument is

$$L_{\lambda}^r(\omega_r) = \frac{\int_0^{\Delta_{\text{FOV}}} \int_{\lambda_1}^{\lambda_2} L_{\lambda}^r(\lambda, \omega) \varrho(\lambda) d\lambda d\omega}{\int_0^{\Delta_{\text{FOV}}} d\omega},\quad (12)$$

where ω is used to represent the spherical coordinate zenith and azimuth angle pair (θ, ϕ) and the integral over solid angle has been abbreviated as

$$\int_0^{\Delta\omega} d\omega = \int_0^{2\pi} \int_0^{\Delta\theta} \sin \theta d\theta d\phi.\quad (13)$$

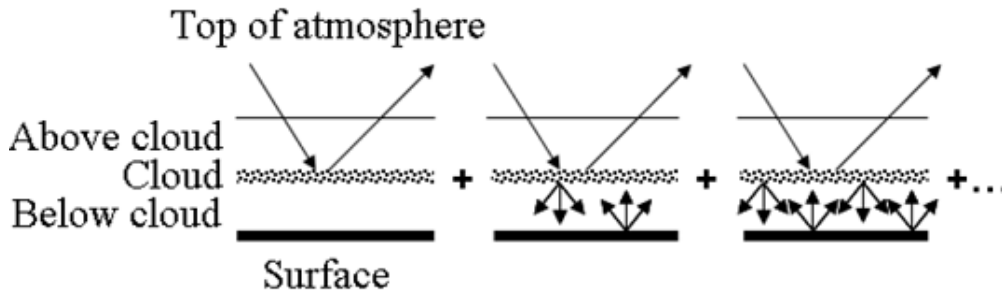


Figure 2: Schematic of the contributions to the measured radiance through multiple scattering between the atmosphere and surface.

The ‘Sun-normalised radiance’ (or top-of-atmosphere reflectance) can then be formed by dividing the measured radiance $L_{\lambda}^r(\omega_r)$ by E_{λ}^0 , the irradiance the satellite would measure if viewing the Sun through a perfect diffuser i.e.

$$R(\bar{\lambda}, \omega_0, \omega_r) = \frac{\pi L_{\lambda}^r(\omega_r)}{\cos \theta_0 E_{\lambda}^0}. \quad (14)$$

The factor $\cos \theta_0$ accounts for the reduction in energy per unit area when the Sun’s energy strikes the atmosphere-Earth system at an angle θ_0 to the local vertical.

In the limit of a very narrow band, the measured Sun normalised radiance is a good approximation to the spectral bidirectional reflectance factor $R(\lambda, \omega_i, \omega_r)$, which is defined as the ratio of the reflected radiant flux to the reflected radiant flux from an ideal diffuse (i.e. Lambertian) surface [27]. The bidirectional reflectance factor is a function of the wavelength λ and the input and output directions (represented by ω_i and ω_r , respectively). For simplicity, the dependence of reflection and transmittance on λ will not be explicitly shown.

The cloudy fraction of the atmosphere in a scene is modelled with three layers: a below-cloud layer, a cloud layer and an above-cloud layer. The above and below-cloud layers consist of gaseous absorbers that attenuate radiation without scattering³.

The gaseous absorption optical depth of the atmosphere is calculated using visible channel coefficients for RTTOV version 12.0 and the clear sky contribution for each scene is calculated with NWP information provided by 6-hourly ECMWF ERA Interim analyses. This total absorption optical depth is then partitioned into the above-cloud optical depth τ_{ac} and the below-cloud optical depth τ_{bc} based on the cloud top pressure relative to the surface pressure.

Using the reflectance and transmission operators described in section 2.4.2, the surface reflectance description in section 2.4.3, and by neglecting molecular absorption, the observed reflectance of the atmo-

³Molecular scattering throughout the atmospheric column is included in the scattering calculations carried out for the cloud layer



sphere/surface system can be written as (assuming dependence on wavelength λ):

$$\begin{aligned}
 R(\theta_0, \theta_v, \phi) = & R_{bb}(\theta_0, \theta_v, \phi) && \text{Reflection off the atmosphere} \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bb}(\theta_0, \theta_v, \phi) T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) T_{db}^{\uparrow}(\theta_v) && \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{Single reflection off the surface} \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} T_{db}^{\uparrow}(\theta_v) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) R_{dd} \rho_{db} T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) R_{dd} \rho_{dd} T_{db}^{\uparrow}(\theta_v) && \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{Double reflection off the surface} \quad (15) \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} R_{dd} \rho_{db} T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} R_{dd} \rho_{dd} T_{db}^{\uparrow}(\theta_v) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) R_{dd} \rho_{dd} R_{dd} \rho_{db} T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) R_{dd} \rho_{dd} R_{dd} \rho_{dd} T_{db}^{\uparrow}(\theta_v) && \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{Triple reflection off the surface} \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} R_{dd} \rho_{dd} R_{dd} \rho_{db} T_{bb}^{\uparrow}(\theta_v) \\
 & + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} R_{dd} \rho_{dd} R_{dd} \rho_{dd} T_{db}^{\uparrow}(\theta_v) \\
 & + \dots
 \end{aligned}$$

Here we have four terms resulting from a single surface reflection in equation 15, which can be described as follows:

- $T_{bb}^{\downarrow}(\theta_0) \rho_{bb}(\theta_0, \theta_v, \phi) T_{bb}^{\uparrow}(\theta_v)$ is the direct transmission of the solar beam, reflected off the surface and transmitted directly to the satellite.
- In $T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) T_{db}^{\uparrow}(\theta_v)$ the diffusely reflected portion of the directly transmitted solar beam is diffusely transmitted (via multiple scattering in the atmosphere) into the viewing direction of the satellite.
- $T_{bd}^{\downarrow}(\theta_0) \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v)$ gives the portion of the diffusely transmitted solar beam, which is then reflected into the viewing direction of the satellite and directly transmitted back through the atmosphere.
- $T_{bd}^{\downarrow}(\theta_0) \rho_{dd} T_{db}^{\uparrow}(\theta_v)$ is the purely diffuse component, where solar radiation is diffusely transmitted to the surface, reflected off the surface and diffusely transmitted to the satellite.

The terms following on from these describe the rapidly diminishing series of multiple reflections between the surface and overlaying atmosphere. For these terms the assumption has been made that ground and atmosphere pair are essentially Lambertian reflectors (i.e. that only the bi-hemispherical reflectance of the atmosphere is needed). Neglecting directly transmitted solar radiation, this is equivalent to saying the sky is equally bright in all directions.

By gathering terms, equation 15 can be simplified to give

$$\begin{aligned}
 R = & R_{bb}(\theta_0, \theta_v, \phi) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bb}(\theta_0, \theta_v, \phi) T_{bb}^{\uparrow}(\theta_v) + T_{bd}^{\downarrow}(\theta_0) \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v) \\
 & + \left(T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} \right) T_{db}^{\uparrow}(\theta_v) (1 + \rho_{dd} R_{dd} + \rho_{dd}^2 R_{dd}^2 + \dots) \\
 & + \left(T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} \right) R_{dd} \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v) (1 + \rho_{dd} R_{dd} + \rho_{dd}^2 R_{dd}^2 + \dots).
 \end{aligned} \quad (16)$$

This can then be further simplified, using the appropriate series limit, to give

$$\begin{aligned}
 R = & R_{bb}(\theta_0, \theta_v, \phi) \\
 & + T_{bb}^{\downarrow}(\theta_0) \rho_{bb}(\theta_0, \theta_v, \phi) T_{bb}^{\uparrow}(\theta_v) + T_{bd}^{\downarrow}(\theta_0) \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v) \\
 & + \frac{\left(T_{bb}^{\downarrow}(\theta_0) \rho_{bd}(\theta_0) + T_{bd}^{\downarrow}(\theta_0) \rho_{dd} \right) \left(T_{db}^{\uparrow}(\theta_v) + R_{dd} \rho_{db}(\theta_v) T_{bb}^{\uparrow}(\theta_v) \right)}{1 - \rho_{dd} R_{dd}}.
 \end{aligned} \quad (17)$$



Finally, the observed TOA reflectance including molecular absorption is obtained by scaling the terms in equation 17 by the appropriate clear-sky transmission terms

$$\begin{aligned}
 R_{\text{TOA}} = & \mathcal{T}_{\text{ac}}(\theta_0)\mathcal{T}_{\text{ac}}(\theta_v) [R_{\text{bb}}(\theta_0, \theta_v, \phi) \\
 & + \mathcal{T}_{\text{bc}}(\theta_0)T_{\text{bb}}^\downarrow(\theta_0)\rho_{\text{bb}}(\theta_0, \theta_v, \phi)T_{\text{bb}}^\uparrow(\theta_v)\mathcal{T}_{\text{bc}}(\theta_v) + \mathcal{T}_{\text{bc}}(0)T_{\text{bd}}^\downarrow(\theta_0)\rho_{\text{db}}(\theta_v)T_{\text{bb}}^\uparrow(\theta_v)\mathcal{T}_{\text{bc}}(\theta_v) \\
 & + \frac{\left(\mathcal{T}_{\text{bc}}(\theta_0)T_{\text{bb}}^\downarrow(\theta_0)\rho_{\text{bd}}(\theta_0) + \mathcal{T}_{\text{bc}}(0)T_{\text{bd}}^\downarrow(\theta_0)\rho_{\text{dd}}\right) \left(\mathcal{T}_{\text{bc}}(0)T_{\text{db}}^\uparrow(\theta_v) + R_{\text{dd}}\mathcal{T}_{\text{bc}}^2(0)\rho_{\text{db}}(\theta_v)T_{\text{bb}}^\uparrow(\theta_v)\mathcal{T}_{\text{bc}}(\theta_v)\right)}{1 - \rho_{\text{dd}}R_{\text{dd}}\mathcal{T}_{\text{bc}}^2(0)}] ,
 \end{aligned} \tag{18}$$

where $\mathcal{T}_{\text{ac}}(\theta) = e^{-\tau_{\text{ac}}/\cos\theta}$, $\mathcal{T}_{\text{bc}}(\theta) = e^{-\tau_{\text{bc}}/\cos\theta}$, τ_{ac} and τ_{bc} are the above cloud and below cloud optical thicknesses, respectively, and it is assumed that the mean angle of diffuse transmission is 66° .

2.4.5 Thermal-IR RTM

The thermal RTM makes extensive use of the RTTOV model [23]. RTTOV directly provides the modelled radiance from the clear-sky fraction of the scene.

The observed cloudy brightness temperature is given by

$$L^\uparrow(\theta_v) = L_{\text{ac}}^\uparrow(\theta_v) + \left(L_{\text{ac}}^\downarrow R_{\text{db}}^\uparrow(\theta_v) + B(T_c)\varepsilon_c + L_{\text{bc}}^\uparrow T_{\text{db}}^\uparrow(\theta_v)\right) e^{-(\tau_{\text{ac}}/\cos\theta_v)}, \tag{19}$$

where $L_{\text{ac}}^\uparrow(\theta_v)$ is the upward radiance into the viewing direction from the atmosphere above the cloud, L_{ac}^\downarrow is the downward radiance from the atmosphere above the cloud, L_{bc}^\uparrow is the upward radiance from the atmosphere below the cloud, $B(T_c)$ is the Planck function as a function of the cloud top temperature T_c , ε_c is the cloud emissivity obtained from an LUT computed in a similar way as those for the operators R and T and $e^{-(\tau_{\text{ac}}/\cos\theta_v)}$ is the transmission from TOA to the cloud top. The transmission term $e^{-(\tau_{\text{ac}}/\cos\theta_v)}$ is obtained from transmission profiles computed with RTTOV while the clear-sky radiance terms L are obtained from thermal emission profiles computed with RTTOV.

2.4.6 Derivatives of the forward model

The gradient of the forward model $\partial y_i/\partial x(j)$, where y_i is a radiance measurement in a single channel and x_j is one of the retrieved parameters, is required for the following purposes:

1. The gradient with respect to parameters which are to be derived from the measurements (state parameters) is a vital quantity for the inversion of the non-linear reflectance model by the Levenberg-Marquardt algorithm.
2. The gradient with respect to parameters which might be considered known and not part of the inversion procedure (model parameters such as surface reflectance spectral shape) is used to judge the sensitivity to these parameters and thus to estimate their contribution to the retrieval uncertainty.

Derivatives of the forward model may be obtained through straightforward linearisation of the forward model equations already given and as a result will not be listed here.

2.4.7 Single layer Measurement vector and covariance

The retrieval scheme described here uses nadir-view observations in the 0.67, 0.87, 1.6, 3.7, 11 and 12 μm channels. In practice only one of the 1.6 or 3.7 μm channels is included in a given retrieval because it has been



found to be difficult to consistently represent both the 1.6 and 3.7 μm channels with the simple CM used [2]. Similarly, forward view radiances are not included as the three-dimensional structure of cloud will often cause differences between the views which cannot be accommodated by the simple model. The error covariance used in the retrieval is the sum of three terms:

$$\mathbf{S}_y = \mathbf{S}_{\text{noise}} + \mathbf{S}_{\text{pixel}} + \mathbf{S}_{\text{fm}}. \quad (20)$$

- $\mathbf{S}_{\text{noise}}$ represents random instrument noise on the observations. The matrix is assumed diagonal with values on the diagonal equal to the square of the assumed measurement noise, which are set for each instrument based on pre-launch characterisation.
- $\mathbf{S}_{\text{pixel}}$ represents errors related to inadequacies of the plane-parallel cloud model and imperfect co-registration of the channels. It is assumed diagonal and equal to the square of 2% of the measured radiance for visible and near-IR channels⁴, and 0.08 K for the thermal channels. These are combined for the mixed 3.7 μm channel⁵. See Watts et al. [36] for the derivation of this term.
- \mathbf{S}_{fm} is zero for rows and columns corresponding to thermal channels. For the visible and near-IR channels, the matrix represents uncertainties in the MODIS surface albedo. Diagonal elements are set to (the square of) 20% of the albedo for the corresponding channel. Off-diagonals are set to give a correlation between the visible/near-IR channels of 0.2.

2.4.8 State vector and a priori constraint

The state-vector used in the retrieval prescribes:

- Log_{10} optical depth;
- Cloud effective radius;
- Cloud top pressure;
- Surface temperature.
- Cloud fraction

A priori and first guess values depend on the assumed phase (phase determination is addressed in section 2.2). For liquid cloud, *a priori* values are 6.3, 12 μm and 900 hPa for optical depth, effective radius and cloud top pressure, respectively; for ice the equivalent values are 6.3, 30 μm and 400 hPa. The surface temperature *a priori* value is taken from ECMWF reanalysis fields for the skin temperature. Currently the cloud fraction is not retrieved as it was found that allowing the cloud fraction to vary introduced compensating effects. However in the future if subpixel cloud fraction was available with improved confidence this could be easily reintroduced. A more complete exploration of the sea surface temperature retrieval can be found in (author?) [Cox et al].

In the absence of useful information, the *a priori* uncertainty for the state parameters is set to 10^8 for optical depth, effective radius, and cloud top pressure. This implies that these parameters are effectively unconstrained by their *a priori* value. For surface temperature, the *a priori* uncertainty is set to 2 K over sea and 5 K over land. The *a priori* covariance is assumed to be diagonal.

The initial (first guess) values for state-vector parameters are set equal to the *a priori* values, except cloud top pressure which is set by:

⁴For ATSR-2 channel 4, the visible uncertainty is 1.5 %.

⁵For mixed channels, the radiance is converted into an equivalent brightness temperature.



- Find the channel with wavenumber less than 2500 cm^{-1} that is nearest to 909 cm^{-1} . If there is no such channel, do not attempt a retrieval.
- Convert the brightness temperature in that channel to a radiance. If the pixel is flagged as clear, subtract the TOA clear-sky radiance (as given by the LUT for that channel).
- If that radiance is less than zero, use the first guess specified above. Otherwise, convert the radiance into a brightness temperature and interpolate it onto the ECMWF temperature profile to find the first guess cloud top pressure.

2.4.9 Cloud Emissivity and microphysical retrievals at night and twilight

The retrieval of cloud top height typically relies on matching thermal emission from a cloud in the $11 \mu\text{m}$ window to a vertical location in a known temperature profile. Without accounting for cloud transparency, this method produces heights that are biased low for semi-transparent clouds due to the contribution of thermal emission from below the cloud. Including additional thermal channels allows for the possibility of retrieving information on cloud transparency from which it is possible to separate the cloud from the below-cloud signal. In particular, the $3.7\text{-}\mu\text{m}$ channel paired with $11 \mu\text{m}$ has been used to retrieve cloud top height along with a single microphysical parameter that describes the cloud radiative thickness in the infrared, typically referred to as the “effective emissivity”. In a similar manor it has been shown that the relative interdependence of optical thickness and effective radius in the 3.7 and $11.0 \mu\text{m}$ combination allows these to be retrieved along with cloud top pressure, although with a significantly greater uncertainty than using solar wavelengths during the day.

The night retrieval is run where the solar zenith angle is greater than 90° or at least two visible/near-IR channels do not convey valid information. Twilight conditions are defined by solar zenith angles between 80° and 90° , where the $3.7 \mu\text{m}$ channel is not used due to solar contamination. Retrievals in Twilight conditions are performed however the quality is poor due to the difficulty in modelling the solar contribution of the $3.7 \mu\text{m}$ channel.

2.4.10 Cloud Albedo

Given the reflection function is parameterised as a function of waveband, input and output geometry and cloud properties (optical depth, effective radius and phase), it is possible to output three reflection terms:

- The cloud reflection factor for the Sun-cloud-satellite geometry $R(\mu_0, \phi_0; \mu, \phi)$;
- The directional-hemispherical reflectance for unidirectional illumination (also called the black-sky albedo)

$$R(\mu_0) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\mu_0, \phi_0; \mu, \phi) \mu \, d\mu \, d\phi; \quad (21)$$

- The cloud bihemispherical reflectance for isotropic illumination (also called the white-sky albedo)

$$R = \frac{1}{\pi^2} \int_0^{2\pi} \int_0^1 \int_0^{2\pi} \int_0^1 R(\mu_0, \phi_0; \mu, \phi) \mu \, d\mu \, d\phi \, \mu_0 \, d\mu_0 \, d\phi_0. \quad (22)$$

Note that these terms are calculated in the absence of knowledge of the underlying surface so they refer to a cloud property alone.

The black-sky albedo is currently the reported cloud albedo in the CCI product.



2.4.11 Corrected CTT, CTP and CTH

The temperature of the geometric top of the cloud, which we have named as the *corrected* cloud top temperature $T_{c,cor}$, in contrast to the retrieved radiative cloud top temperature T_c (CTT), may be written as

$$T_{c,cor} = BT_c(\lambda) + \Gamma \Delta Z, \quad (23)$$

where $BT_c(\lambda)$ is the brightness temperature (K) at wavelength λ observed from the *top of the cloud*, Γ is the atmospheric lapse rate (-K/km) and ΔZ (km) is the geometric depth into the cloud of T_c .

If we assume that the cloud is vertically homogeneous the cloud optical depth $\Delta\tau_c(\lambda)$ at a geometric depth ΔZ (km) into the cloud is given by

$$\Delta\tau_c(\lambda) = \sigma_c(\lambda)N_c\Delta Z, \quad (24)$$

where $\sigma_c(\lambda)$ is the size distribution averaged cloud particle cross section and N is the cloud particle number concentration ($1/\text{km}^3$). If we assume that the peak sensitivity to T_c comes from one optical depth into the cloud ($\Delta\tau_c = 1$) then ΔZ may be written as

$$\Delta Z = \frac{1}{\sigma_c(\lambda)N_c}. \quad (25)$$

From equation 23, two equations may be formed, for measurements at 11 and 12 , respectively:

$$T_{c,cor} = BT_c(11) + \frac{\Gamma}{\sigma_c(11)N_c}, \quad (26)$$

$$T_{c,cor} = BT_c(12) + \frac{\Gamma}{\sigma_c(12)N_c}, \quad (27)$$

where there are two unknowns, $T_{c,cor}$ and Γ/N_c . By solving for Γ/N_c with each equation and equating the results as

$$\sigma_c(11) [T_{c,cor} - BT_c(11)] = \sigma_c(12) [T_{c,cor} - BT_c(12)], \quad (28)$$

$T_{c,cor}$ may be solved for, given by

$$T_{c,cor} = \frac{\sigma_c(11)BT_c(11) - \sigma_c(12)BT_c(11)}{\sigma_c(11) - \sigma_c(12)}. \quad (29)$$

The brightness temperatures as observed from the top of the cloud, $BT_c(\lambda)$ may be obtained from the corresponding brightness temperatures observed at TOA $BT_{TOA}(\lambda)$ with

$$BT_c(\lambda) = BT_{TOA}(\lambda)/\mathcal{T}_{ab}(\lambda), \quad (30)$$

where $\mathcal{T}_{ab}(\lambda)$ is the above cloud clear-sky atmospheric transmittance, i.e. the transmittance from TOA to the top of the cloud, along the satellite viewing slant path. Cloud particle cross section $\sigma_c(\lambda)$ is obtained from pre-computed LUTs as a function of wavelength and the retrieved cloud optical thickness (COT) and cloud effective radius (CER).

The corresponding corrected cloud top pressure $P_{c,cor}$ and corrected cloud top height $H_{c,cor}$ may be obtained from the input pressure and temperature profiles simply by comparing $T_{c,cor}$ with the input temperature profile and interpolating appropriately.

The estimated uncertainties of $T_{c,cor}$, $P_{c,cor}$ and $H_{c,cor}$ are obtained through straight forward error propagation, the details of which will not be discussed here.



2.4.12 Boundary layer inversion

A boundary layer temperature inversion can often permit two valid solutions for the cloud top pressure within the OEM retrieval, depending on whether one searches the temperature profile from top to bottom or vice versa. Thus, the retrieved height may either be too high or too low. The following solution is derived from the Meteosat Third Generation Level 2 Processing Specification Document [17] but is repeated here as that document is not easily accessible:

- Search from the bottom of the forecast temperature profile for a typical anticyclonic temperature inversion. Numbering the levels of the forecast model from 1 at the surface to N at TOA, an inversion is identified by the smallest $1 < i < N$ where $T_{\text{fcst}}^i < T_{\text{fcst}}^{i+1} - 1 \text{ K}$ and $T_{\text{fcst}}^i < T_{\text{fcst}}^{i-1}$ with $P_{\text{fcst}}^i > 600 \text{ hPa}$ (where T_{fcst}^i and P_{fcst}^i are the temperature and pressure at forecast level i , respectively). Eventually, this scheme should be extended to account for the humidity profile.
- The top of the inversion is identified as the smallest $j \geq i + 2$ at which $T_{\text{fcst}}^j < T_{\text{fcst}}^{j-1}$.
- If no inversion is found, the forecast temperature profile is retained. Otherwise, modify the temperature profile within the inversion to facilitate a more sensible cloud top pressure retrieval. This extrapolates the boundary layer below the inversion such that:

$$T^k = \begin{cases} T_{\text{fcst}}^i + \Gamma_{\text{BL}} (P_{\text{fcst}}^k - P_{\text{fcst}}^i) & i < k \leq j + 2 \\ T_{\text{fcst}}^k & \text{otherwise,} \end{cases} \quad (31)$$

where the boundary layer lapse rate Γ_{BL} is approximated from the levels beneath the inversion,

$$\Gamma_{\text{BL}} = \frac{dT}{dP} \simeq \frac{T_{\text{fcst}}^{i-1} - T_{\text{fcst}}^{i-2}}{P_{\text{fcst}}^{i-1} - P_{\text{fcst}}^{i-2}}. \quad (32)$$

- The +2 in eqn. 31 and the second step account for uncertainty in the bounding of the inversion due to forecast model errors and smoothing.

2.4.13 Tropopause identification and temperature profile modification

Occasionally deep convective clouds rise above the tropopause and there is a delay before this is expressed in the temperature profile. If not corrected, the retrieved cloud top height will be too low. This is accounted for, equivalent to Section 2.4.12, by extrapolating from the temperature profile just below the tropopause:

- The tropopause is identified by the largest $i < N$ where $T^i > T^{i+1}$, $T^{i-2} - T^{i-1} > 2 \text{ K}$ and $P_{\text{fcst}}^{i+1} > 80 \text{ hPa}$.
- The temperature profile is then modified such that:

$$T_{\text{modified}}^k = \begin{cases} T^i + \Gamma_{\text{troposphere}} (P_{\text{fcst}}^k - P_{\text{fcst}}^i) & i < k \leq N \\ T^k & \text{otherwise,} \end{cases} \quad (33)$$

where the tropospheric lapse rate $\Gamma_{\text{troposphere}}$ is approximated from the levels beneath the tropopause,

$$\Gamma_{\text{troposphere}} = \frac{T_{\text{fcst}}^{i-2} - T_{\text{fcst}}^i}{P_{\text{fcst}}^{i-2} - P_{\text{fcst}}^i}. \quad (34)$$



2.4.14 ECMWF data

Clear-sky atmospheric radiances and transmittances are determined by RTTOV. This requires meteorological information as an input, which is provided by ECMWF ERA Interim reanalysis fields [7]. The required data⁶ are stored in one or more files of NetCDF or GRIB format. The data presents values and profiles representative of the atmosphere and surface at each point on a reduced Gaussian grid (with resolution dependent on the data source). ORAC retrieves states averaged over each satellite pixel and, to reduce computational expense, only evaluates RTTOV on a regular 500-by-500 latitude-longitude grid, which is then linearly interpolated onto each satellite pixel. (The errors from this process have been found to be less than the uncertainties in the RTTOV calculations themselves and so are considered negligible.)

The ECMWF data must be interpolated onto the ORAC grid. As the ECMWF data is generally on a coarser grid than that used by ORAC, the grid-cell average is approximated by the ECMWF value at the cell centre. The interpolation is performed using the EMOSLIB library provided by ECMWF (found at software.ecmwf.int/wiki/display/EMOS/Emoslib).

2.4.15 Quality Control

The quality of the ORAC algorithm is based on the output diagnostics of the OE retrieval.

- The number of iterations: Indicates if the retrieval has converged.
- A convergence test: ORAC uses the change in the cost function between iterations to determine whether a retrieval has converged.
- Cost function: If the cost function is approximately equal to the number of measurements then the retrieval is thought to have fit the model well. In practice, any retrieval with cost greater than ten times the number of measurements is considered suspect.
- Error estimates: If the previous criteria is satisfied then the uncertainty on the retrieved parameters is given in the solution covariance.

2.4.16 Modifications for ATSR-2/AATSR/SLSTR

After processing version 2.0 of the ATSR-2/AATSR climatology a significant inconsistency between the two sensors was discovered. This was primarily due to the 'narrow' swath geometry of ATSR-2. A correction was made to identify when ATSR-2 was in narrow swath mode and only process in the full swath. The version 3 of the ATSR2-AATSR record includes this implementation. In addition, version 3 was processed using the 1.6 μ m channel instead of the 3.7 μ m channel.

2.4.17 Aerosol-Cloud (and sunglint) discrimination

Evaluation have shown that occasionally heavy aerosol loading are misclassified as clouds. Two empirical reclassification approaches have been developed and applied.

1. Removal of clouds over the Sahara region if: (i) the total `cldmask_uncertainty` (which is the sum of both views when used from AATSR) is greater then 70 and (ii) the Pavolonis cloud type is water (in either view in case of AATSR two views retrievals).

⁶Currently, the fields `temperature`, `spec_hum`, `ozone`, `geopot`, `lnsp`, `sea_ice_cover`, `snow_albedo`, `sst`, `totcolwv`, `snow_depth`, `u10`, `v10`, `temp2`, `land_sea_mask` and `skin_temp` are required.



2. Removal of clouds based on combined testing of (i) the $11\mu\text{m}$ - $12\mu\text{m}$ brightness temperature difference (being between 0 and 3K), (ii) the Brightness temperature in $11\mu\text{m}$ being larger than 290K and (iii) the spatial (in-)homogeneity in the $11\mu\text{m}$ channel (standard deviation in 3×3 pixels) being smaller 0.25K. If all of these requirements are fulfilled the pixel is reclassified as clear-sky in case it was classified as cloudy beforehand. The cloud mask uncertainty is set to 99% in such pixels.

2.5 Retrieval Performance

For a comprehensive evaluation of the retrieval performance the reader is referred to Mcgarragh et al. [15] (**author?**) [PVIR] and [31]. The results are briefly summarised here. The performance of the retrieval method was assessed theoretically by simulating measurements using a range of values for the retrieval parameters and then subsequently performing a retrieval on these simulated measurements to which Gaussian noise levels appropriate for MODIS were added. The errors are less than 10% could have implications for the retrieval of subvisible cirrus. When the algorithm is run in 'heritage' mode the cloud top height for multi layer cloud systems with a thin (less than 4 COT) is located between the upper and lower layer at the radiative cloud top height. The cloud fraction has been validated for different cloud regimes and found to perform well over sea and target surfaces with performance reduced over the polar regions.

2.6 Changes compared to precursor CC4CL version

Between ATBDv5.1 (describing CC4CL as used for dataset versions 2) and current version (used for dataset version 3) a number of changes and significant bug fixes were made to the retrieval algorithm, which are listed in the following.

Changes applying to the processing of both AVHRR and ATSR2-AATSR:

- Retraining ANN for cloud detection incorporating a much larger data basis and with the emphasis on NOAA-19 AVHRR. Furthermore, new channel settings for ANN input, new scene dependent thresholds for deriving binary cloud mask, new spectral band adjustment to make all sensors mimic NOAA-19 AVHRR.
- Introduction of a ANN for phase identification with a very similar approach as for cloud detection, also using CALIOP phase as truth. By this cloud phase uncertainty is introduced as well, derived in a similar manner as for cloud detection. Cloud typing remains to be based on Pavlonis schemes.
- Implementing BRDF calculation bug fix, in particular improving retrievals in polar regions
- Using a spatially higher resolved land/sea mask dataset (USGS 0.5km MODIS based Land Cover Type 2 data)
- A comprehensive review of ice optical properties was undertaken and a new set of forward model LUTs based on Baum [37] optical properties were derived that show improved results, most significantly for cloud ice effective radius (and ice water path).
- New QC information regards sunglint and surface type (e.g. ice and snow), and includes a simplified high quality/low quality flag meant for general users.
- Including BUGSRAD scheme for calculation of radiative broadband fluxes at top of atmosphere and surface (bottom of atmosphere). See CC4CL TOA FLUX ATBD (available at <http://www.esa-cloud-cci.org/?q=documentation>) for details.
- Implemented time dependent CO_2 in RTTOV and BUGSRAD calculations.



Changes applying to processing of ATSR2-AATSR

- Changed processing of cloud properties when in narrow swath mode.
- Processed using 1.6 μ m channel rather than 3.7 μ m channel which has significant dynamic range
- Fixed bug due to overlapping orbits
- Used NISE ice/snow mask which have higher resolution and cover ATSR time period.

3 Input and output data

The primary input data used by CC4CL are calibrated, geolocated satellite radiances, generally referred to as level 1 data.

- When applied to (A)ATSR data, CC4CL ingests standard ESA level 1b data.
- NASA M*D021KM radiance and M*D03 geolocation files are the primary input for MODIS retrievals, where the “*” represents either an “O” for MODIS-TERRA or “Y” for MODIS-AQUA data.
- AVHRR data are ingested via the level 1 “avhrr” and “sunsatangles” files obtained from “pygac”, which is a python package to read and calibrate NOAA AVHRR GAC data. It has been developed by SMHI (Swedish Meteorological and Hydrological Institute) as a standalone tool based on the precursor “AHAMAP”, which was used in the PPS (Polar Platform System) 2010 CLARA-A1 processing in CMSAF. DWD (Deutscher Wetterdienst) has reprocessed AVHRR level 1 data using pygac in order to provide the required input for Cloud_cci.

The physical quantity measured by satellite radiometers is radiance. Two modifications are made to the radiance for use with CC4CL: the shortwave channels are scaled by the cosine of the solar zenith angle and normalised, to produce a sun-normalised reflectance, and the thermal channel radiances are converted to brightness temperatures in Kelvin.

These files then provide CC4CL with:

- Calibrated TOA reflectance/brightness temperature;
- Solar and satellite azimuth and zenith angles;
- A land/sea mask for the level 1 grid.

CC4CL also makes use of a range of ancillary data:

- ECMWF ERA Interim humidity and temperature profiles, total column ozone, surface temperature and pressure, and 10 m east-west and north-south (u and v) wind components. These are used by the sea surface reflectance model to determine surface roughness and whitecap coverage and to estimate the sea surface emissivity.
- to ensure historical consistency ECMWF snow/mask can be used, while an option has been implemented to use higher resolution snow masks (e.g. NISE). The snow/ice information is used to modify the surface albedo and BRDF calculations.
- MODIS MCD43C1 surface BRDF product produced every 8 days with a 16-day acquisition from both Aqua and Terra-based MODIS observations [12, 26].



- The emissivity over land is taken from the CIMSS database [28]. Optional is the utilization of an ancillary land/sea mask, overwriting land/sea information contained in L1 data. This option has recently been updated to make use spatially highly resolved USGS 0.5km MODIS based Land Cover Type 2 data (https://archive.usgs.gov/archive/sites/landcover.usgs.gov/global_climatology.html)

The parameters retrieved by CC4CL are constrained to the following ranges:

- The \log_{10} of cloud optical depth at $0.55 \mu\text{m}$: $[-3 - 2.408]$
- The effective radius (in μm): $[0.1 - 35 \text{ (liquid) or } 100 \text{ (ice)}]$
- The cloud top pressure (in hPa): $[10 - 1200]$
- Surface temperature (in K): $[250 - 320]$

In addition to the retrieved state parameters, a number of cloud variables are derived and stored in the global-level cloud products. Cloud top pressure is converted to cloud top height and temperature using the ECMWF temperature/pressure profiles. Cloud water path CWP is derived using the method of Han et al. [8]

$$CWP = \frac{4}{3} \frac{\tau \cdot r_{\text{eff}} \cdot \rho}{Q_{\text{ext}}}, \quad (35)$$

where Q_{ext} , the extinction coefficient, is assumed to be 2 for water and 2.1 for ice for wavelengths much less than r_{eff} . The density ρ is 1 g m^{-3} for water and 0.9167 g m^{-3} for ice. Depending on phase, CWP is also known as liquid water path (LWP) or ice water path (IWP).

More technical descriptions of CC4CL's inputs, outputs, and data formats, in addition to example data files, can be found at the project's code repository at <https://github.com/ORAC-CC/orac>.



Appendices

A Bayesian scene identification

Although not part of the operational CC4CL scheme used to produce the Phase I Cloud_cci products, the use of Bayesian statistics on the retrieval output to refine the cloud masking and phase selection – together described as scene identification – has been investigated and implemented as an option in the retrieval algorithm. A brief description of the methodology and initial results is provided here.

A posteriori Bayesian scene identification is concerned with answering the question, “What is the probability that a given retrieval result is consistent with the measurements and our prior knowledge of the atmospheric state?” Given an answer to that question for a range of assumed cloud and/or aerosol properties, we can further ask whether the measurements provide a clear indication of which set of those assumptions best describes the observations.

As described in section 2, an OEM retrieval is constructed based on maximising the Bayesian probability of the retrieved state, conditional on the measurements and *a priori* constraints. This formulation assumes that the uncertainties in both the measurement and *a priori* are Gaussian and that the retrieval is nearly linear within the region defined by these uncertainties. Rodgers [22] shows that the cost function $J(\vec{x})$ is directly related to this conditional probability through the expression

$$P(\vec{x}|\vec{y}, \vec{x}_a) = P(\vec{x}_a) \exp\left(-\frac{J(\vec{x})}{2}\right). \quad (36)$$

It is clear, therefore, that a well-characterised optimal estimation retrieval scheme not only provides state estimates with fully propagated uncertainties but can also provide probabilistic measures of the consistency of the state with the measurement and *a priori*.

If we examine the expression for the OEM cost function:

$$J(\vec{x}) = [\vec{y}(\vec{x}) - \vec{y}_m] \mathbf{S}_y^{-1} [\vec{y}(\vec{x}) - \vec{y}_m]^T + (\vec{x} - \vec{x}_a) \mathbf{S}_a^{-1} (\vec{x} - \vec{x}_a)^T, \quad (37)$$

we can identify the two terms on the right hand side: the measurement cost J_m and the *a priori* cost J_a . In situations where the forward model is a good representation of the true atmosphere, in that it is able to reproduce the observed TOA radiances while remaining consistent with the prior constraints, these terms should represent random samples from normal distributions with differing degrees of freedom. J_a is the degrees of freedom equal the number of elements in the state vector. J_m is equal to the degrees for freedom of signal d_s , which is the number of independent pieces of information provided by the measurement above the level of noise and is defined by:

$$d_s = \text{tr}\left(\mathbf{K}\mathbf{S}_a\mathbf{K}^T [\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_y]^{-1}\right), \quad (38)$$

where $\text{tr}(\mathbf{A})$ denotes the trace of the matrix \mathbf{A} .

Thus, by comparing the cost function components to the χ^2 distribution — the distribution of the sum of the squares of a set of normally distributed random variables — with the appropriate degrees of freedom, we can calculate the probability that any given retrieval lies within the expected range described by the measurement and *a priori* uncertainties. This is the standard χ^2 test and provides the probabilistic measure of consistency for each retrieval. The test can be performed separately for both the measurement and *a priori* constraint and the resulting probabilities multiplied to produce an overall value, denoted by \mathcal{P} .

In the case where the cloud phase is ambiguous (see section 2.2), the \mathcal{P} statistic can provide useful additional insight. Selecting the phase with the highest \mathcal{P} value is equivalent to selecting that with the lowest cost, but in addition one can ascertain the following:



- The probability that each phase is the correct one, based on the measurement and *a priori* constraints. If no phase provides a result which is consistent with the measurement then the atmospheric state lies outside the range of assumptions made in the forward model and the results should be treated with caution.
- Under the assumption that only one phase can be the correct one, we assume that the one with the highest probability is the correct one. In general terms, if N phases have been tried in the retrieval, then the probability that the n th phase is the correct one can be written:

$$\mathcal{P}'_n = \frac{\mathcal{P}_n}{\sum_{i=1}^N \mathcal{P}_i}. \quad (39)$$

The question of phase discrimination can also be generalised to include cloud detection, through the inclusion of background aerosol types (such as maritime aerosol over the oceans) and episodic aerosols, such as desert dust, which might be detected as cloud by traditional threshold-based cloud flagging.

As the \mathcal{P} statistic represents the Bayesian probability of a particular state being consistent with the measurements and prior constraints, within the framework of the retrieval forward model and uncertainty budget, it offers two key advantages:

1. The probability \mathcal{P} represents the true constraint on the atmospheric state provided by the measurement, under the assumptions made in the retrieval. Cloud retrievals are generally under-constrained and a probability provides a intuitive and easily understood measure of the ambiguity of a retrieval result.
2. The availability of a probabilistic cloud mask and phase selection provides a quantitative method of combining the constraint provided by the retrieval itself with prior and posterior constraints through Bayesian methods.

The accuracy — in terms of the probability values produced — of this approach does require that the retrieval uncertainty budget be well understood. If measurement and forward modelling uncertainties are inaccurate, the resulting PDF of the retrieval cost function will not follow the χ^2 distribution and the values of \mathcal{P} produced will be erroneous.

References

- [1] Baran, A. J. and Havemann, S., The dependence of retrieved cirrus ice-crystal effective dimension on assumed ice-crystal geometry and size-distribution function at solar wavelengths, Q. J. R. Meteorol. Soc., 130, 2153–2167, doi: 10.1256/qj.03.154, 2004.
 - [2] Baran A. J., Shcherbakov V.N., Baker B. A., Gayet J. F. and Lawson RP., On the scattering phase function of non-symmetric ice crystals, Q. J. R. Meteor Soc., 131, 260916, 2005.
 - [3] Birks, A., Improvements to the AATSR IPF relating to land surface temperature, ESA Technical note. AATSR Product Handbook, Issue 1.2, European Space Agency, 2004.
- [Cox et al]
- [5] Cox, C. and Munk, W., Measurement of the roughness of the sea surface from photographs on the Sun's glitter, J. Opt. Soc. Am., 44, 838–850, doi:10.1364/JOSA.44.000838, 1954.
 - [6] Cox, C. and Munk, W., Statistics of the sea surface derived from Sun glitter, J. Mar. Res., 13, 198–227, 1954.



- [7] Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F., The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q.J.R. Meteorol. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- [8] Han, Q., Rossow, W. B. and Lasis, A. A., Near-Global Survey of Effective Droplet Radii in Liquid Water Clouds Using ISCCP Data, *J. of Climate*, 7, 465–497, 1994.
- [9] Koepke, P., Effective Reflectance of Oceanic Whitecaps, *AO*, 23, 1816–1824, doi:10.1364/AO.23.001816, 1984.
- [10] Lean, K., Empirical methods for detecting atmospheric aerosol events from satellite measurements, MPhys. Project Report, University of Oxford, http://www.atm.ox.ac.uk/group/eodg/mphys_reports/2009_Lean.pdf, 2009.
- [11] Levenberg, K., A method for the solution of certain nonlinear problems in least squares, *Quart. Appl. Math.*, 2, 164–168, 1944.
- [12] Lucht, W., Schaaf, C.B., and Strahler, A.H., An Algorithm for the Retrieval of Albedo from Space Using Semiempirical BRDF Models, *Atmospheric Science, Surface Radiative Properties/BRDFs, TGRS*, 38, 977–998, doi:10.1109/36.841980, 2000.
- [13] Nolin, A., Armstrong, R. L. and Maslanik, J., Near-Real-Time SSM/I-SSMIS EASE-Grid Daily Global Ice Concentration and Snow Extent. Boulder, Colorado USA: National Snow and Ice Data Center, 1998.
- [14] Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, *SIAM, J. Appl. Math.*, 11, 431–441, doi:10.2307/2098941, 1963.
- [15] 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 2: The optimal estimation approach, *Atmos. Meas. Tech.*, 11, 3397–3431, <https://doi.org/10.5194/amt-11-3397-2018>, 2018.
- [16] Morel, A. and Prieur, L., Analysis of Variations in Ocean Color, *Limnology and Oceanography*, 22, 709–722, doi:10.4319/lo.1977.22.4.0709, 1977.
- [17] MTG Level 2 Processing Specification Document, Document Number EUM/MTG/SPE/12/ 1078, March 2014.
- [18] Pavolonis, M. J. and Heidinger A. K., Daytime Cloud Overlap Detection from AVHRR and VIIRS, *J. Appl. Meteor.*, 43, 762–778, doi:10.1175/2099.1, 2004.
- [19] Pavolonis, M. J., Heidinger A. K., and Uttal, T., Daytime Global Cloud Typing from AVHRR and VIIRS: Algorithm Description, Validation, and Comparisons, *J. Appl. Meteor.*, 44, 804–826, doi:10.1175/JAM2236.1, 2005.
- [20] 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, doi:10.5194/amt-5-1889-2012, 2012.



- [PVIR] Product validation and inter comparison report, http://www.esa-cloud-cci.org/sites/default/files/documents/public/Cloud_cci_D4.1_PVIR_v5.1.pdf
- [22] Rodgers, C. D., Inverse methods for atmospheric sounding: Theory and Practice. Series on Atmospheric, Oceanic and Planetary Physics, Vol. 2, World Scientific, 2000.
- [23] Saunders, R. W., Matricardi, M. and Brunel, P., An Improved Fast Radiative Transfer Model for Assimilation of Satellite Radiance Observations, Q. J. R. Meteor. Soc., 125, 1407–1425, 1999.
- [24] Sayer A. M, Poulsen, C. A., Arnold, C., Campmany, E., Dean, S., Ewen, G. B. L., Grainger, R. G., Lawrence, B. N., Siddans, R., Thomas, G. E., and Watts, P. D., Global retrieval of ATSR cloud parameters and evaluation (GRAPE): dataset assessment Atmos. Chem. Phys., 11, 3913–3936, doi:10.5194/acp-11-3913-2011, 2011.
- [25] Sayer, A. M., Thomas, G. E. and Grainger, R. G., A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, Atmos. Meas. Tech., 2010, 3, 813–838, 4, 2010
- [26] Schaaf, C. B., F. Gao, A. H. Strahler, W. Lucht, X. Li, T. Tsang, N. C. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, P. Lewis, M. Barnsley, P. Hobson, M. Disney, G. Roberts, M. Dunderdale, C. Doll, R. d’Entremont, B. Hu, S. Liang, and J. L. Privette, and D. P. Roy., First Operational BRDF, Albedo and Nadir Reflectance Products from MODIS, Remote Sens. Environ., 83, 135–148, 2002.
- [27] Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., Martonchik J. V., Reflectance quantities in optical remote sensing — definitions and case studies., Remote Sens. Environ., 103, 27–42, 2006.
- [28] Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R. and Huang, H.-L. Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multi-spectral Satellite Radiance Measurements., Appl. Meteor. Climatol., 47, 108–123, 2007.
- [29] R. Siddans, C. Poulsen, E. Carboni Cloud Model for Operational Retrievals from MSG SEVIRI Final Report DRAFT Version 1.0, Eumetsat Contract EUM/CO/07/4600000463/PDW, January 26, 2011
- [30] 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.
- [31] Sus, O., Stengel, M., Stapelberg, S., McGarragh, G., Poulsen, C., Povey, A. C., Schlundt, C., Thomas, G., Christensen, M., Proud, S., Jerg, M., Grainger, R., and Hollmann, R.: The Community Cloud retrieval for CLimate (CC4CL) – Part 1: A framework applied to multiple satellite imaging sensors, Atmos. Meas. Tech., 11, 3373-3396, <https://doi.org/10.5194/amt-11-3373-2018>, 2018
- [32] Stamnes K., Tsay S.C., Wiscombe W. and Jayaweera K., Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media., Appl. Opt., 12, 2502–2509, 1988.
- [33] Takano, Y., and Liou, K. N., Solar radiative transfer in cirrus clouds. Part I: Single-scattering and optical properties of hexagonal ice crystals., J. Atmos. Sci., 46, 3–19, 1989.
- [34] Thomas, G. E., Carboni, E., Sayer, A. M., Poulsen, C. A., Siddans, R. and Grainger, R. G., Oxford-RAL Aerosol and Cloud (ORAC): aerosol retrievals from satellite radiometers, in *Satellite aerosol remote sensing over land*, Kokhanovsky, A. A. and de Leeuw, G. (eds.), Springer-Praxis, 2009.



- [35] Wanner, W., Strahler, A. H., Hu, B., Lewis, P., Muller, J.-P., Li, X., Barker Schaaf, C. L., and Barnsley, M. J., Global Retrieval of Bidirectional Reflectance and Albedo Over Land from EOS MODIS and MISR Data: Theory and Algorithm, *JGR*, 102, D14, 17143–17161, doi:10.1029/96JD03295, 1997.
- [36] Watts, P. D., Mutlow, C. T., Baran, A. J. and Zavody, A. M., Study on cloud properties derived from Meteosat Second Generation Observations. Eumetsat Report, <http://www.eumetsat.int/Home/index.htm>, 1998
- [37] Yang, P., L. Bi, B.A. Baum, K. Liou, G.W. Kattawar, M.I. Mishchenko, and B. Cole, 2013: Spectrally Consistent Scattering, Absorption, and Polarization Properties of Atmospheric Ice Crystals at Wavelengths from 0.2 to 100 μ m. *J. Atmos. Sci.*, 70, 330–347, <https://doi.org/10.1175/JAS-D-12-039.1>
- [38] Zhang Z., Yang, P., Kattawar, G., Riedi, J., Labonnote, L. C., Baum, B. A., Platnick, S., and Huang, H.-L., Influence of ice particle model on satellite ice cloud retrieval: lessons learned from MODIS and POLDER cloud product comparison, *Atmos. Chem. Phys.*, 9, 7115–7129, 2009.