# The Heidelberg Iterative Cloud Retrieval Utilities (HICRU)

# Algorithm Document

# SCIAMACHY Release 1.0

**The Heidelberg Iterative Cloud Retrieval Utilities retrieve effective cloud fraction using the Polarization Monitoring Devices of the SCIAMACHY instrument. Effective cloud fraction means, that one parameter is retrieved, which is sensitive to cloud coverage and -as well- to cloud albedo. The results have to be interpreted as a cloud coverage with respect to a high cloud albedo. Future releases of HICRU will retrieve further cloud parameters. HICRU cloud data is also available for the GOME instrument. This document describes the SCIAMACHY algorithm only.**

**More extended documentation is available on the website: http://satellite.iup.uni-heidelberg.de. Chose "Data Products" and "Cloud data" there.**

## Forward Model

The HICRU algorithm does not use a forward model, but retrieves effective cloud fraction using the threshold method.

## Inversion Procedure

### Summary

The cloud fraction is retrieved through linear interpolation between lower thresholds, which represent clear-sky top of atmosphere radiances, and upper thresholds, which represent top-of-atmosphere radiances of completely cloudy pixels with high cloud albedo.
The thresholds are calculated iteratively dependent on different parameters. The upper thresholds are retrieved as a function of solar zenith angle and the scan angle of the SCIAMACHY instrument. The lower thresholds are calculated iteratively as maps (as a function of the latitude and the longitude) to take the surface albedo into account. Seasonal variations of the albedo are taken into account. The time period used for the retrieval of the lower threshold is selected separately for each day dependent on the availability of measurements detected as cloud free by HICRU.

### Threshold method

The HICRU algorithm is based on the widely used threshold method. First, lower thresholds $I_{cloudfree}$, representing the intensity of cloud free pixels, and upper thresholds $I_{cloudy}$, representing the intensity of completely cloudy pixels, are calculated. The cloud fraction CF is retrieved from the measured intensity $I_{meas}$ through linear interpolation between the thresholds:

$$CF = \frac{I_{meas} - I_{cloudfree}}{I_{cloudy} - I_{cloudfree}}$$

This interpolation assumes a cloud to be lambertian reflector and a SCIAMACHY pixel that can be divided into a cloud free part and a cloudy part, where the albedo of the cloudy part is implicitly determined by the upper thresholds. HICRU uses earthshine radiances divided by the cosine of the solar zenith angle and the daily solar spectrum. This is a first order correction of the dependency of the intensities on the solar zenith angle.

The accuracy of PMD cloud algorithms critically depend on the quality of the calculated lower and upper thresholds. The accurate retrieval of the the lower thresholds are especially important for the detection of cloud free pixels, because the measured intensity is not only sensitive to the cloud coverage and the cloud albedo, but also to the surface albedo, which depends on surface type and the season of the measurement. The lower thresholds are mainly determined by the surface albedo, but also include Rayleigh scattering. PMD algorithms therefore distinguish cloud free and cloudy pixels through intercomparison between cloudy and clear-sky top-of-atmosphere radiances. The major advantage of the HICRU algorithm is the improvement of cloud retrieval through an iterative retrieval of thresholds, including image sequence analysis for the retrieval of the lower thresholds. The algorithms for the retrieval of thresholds are described in in detail below and make an accurate cloud retrieval also possible for regions like deserts, which often cause to problems for SCIAMACHY cloud algorithms.

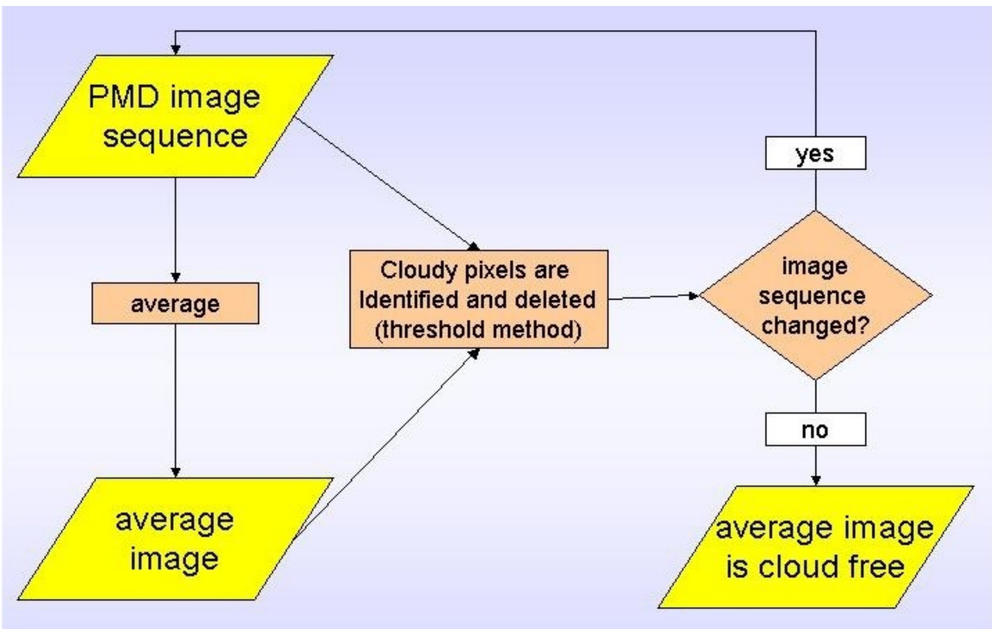
## Instruments used for cloud retrieval

The HICRU algorithm uses the Polarization Monitoring Devices (PMD) of SCIAMACHY to retrieve the effective cloud fraction, because of their higher spatial resolution compared to the channels with moderate spectral resolution. An important advantage of the higher spatial resolution is founded in the strong influence of the surface albedo on the retrieved cloud fraction. The determination of surface albedo requires an adequately large set of measurements referring to cloud free scenarios, but the probability of a cloud free measurement strongly depends on spatial resolution.
There are seven PMD instruments available for SCIAMACHY (PMD1: 310-777nm; PMD2: 450 – 525nm, PMD3: 617 – 705nm, PMD4: 805 – 900nm, PMD5: 1508 – 1645nm, PMD6: 2290 – 2405nm, PMD7: 802 – 905nm). The algorithm is applied to different PMD channels for testing purposes, but for the official data product PMD3 is used:
Some PMDs are not useful for cloud detection (see e.g.http://www.iup.physik.uni-bremen.de/sciamachy/LTM/PMD_LTM/PMD_LTM.html): PMD7 is unusable due to straylight prroblems, PMD4 shows strong degradation effects and PMD6 show a strange variation of the dark signal. PMD1 is omitted by HICRU because of instrument degradation. PMD1 and PMD2 is also not used by HICRU because of the strong impact of Rayleigh scattering and the sensitivity to the polarization of the earth radiance. For the near infrared region covered by PMD5 and PMD6 the assumption used by HICRU, that the albedo of the cloud is higher then the surface is not always appropriate (depending on the surface type). Nevertheless, future releases of HICRU may use the signal of PMD5 and perhaps also the signals of PMD2 and PMD4 additionally (e.g to receive cloud information in the case of snow and ice coverage and for the improvement of the cloud correction of trace gases evaluated in the near-IR wavelength region). But we expect, that the main data product will be still retrieved from PMD3 in the future, because the data product from PMD3 is most robust with respect to the instruments characteristics and the assumption used for the threshold method.

## Thresholds for cloud free measurements (lower thresholds)

The lower threshold strongly depends on surface albedo and has to be calculated with respect to the latitude and longitude of measurement. We have to retrieve a map of the earth containing minimum reflectances of PMD3 as lower thresholds. The spatial resolution (latitude x longitude) of the maps is about 0.100 degree x 0.062 degrees. The resolution is chosen higher in horizontal direction, because the spatial resolution of the PMD measurements is also higher in horizontal direction. Two different strategies could be applied: on the one hand, we should use short periods of time for the retrieval, because of seasonal variations of the surface albedo and the effects of irregular instrument degradation dependent on the time of measurement. Hence we should aim to retrieve maps representing the lower threshold separately for each day using periods as short as possible (HICRU uses 37 days). On the other hand, this method only would work appropriately if cloud free pixels exist during the considered period of time. This assumption holds well for regions like the Sahara, but is hardly acceptable for regions with persistent or seasonal cloud coverage. Note, that SCIAMACHY needs six days to cover the earth completely, thus there are not more than 7 measurements during 37 days for some regions on earth and the possibility that all of them are cloudy is not negligible. The number of available measurements could be even lower than 7, if the level-1-dataset is incomplete for the considered time period. To take both strategies into account, HICRU uses a three stage classification scheme analyzing both long and short periods of time (see below) .

HICRU uses an iterative algorithm based on image sequence analysis for all three stages of threshold retrieval. The main idea is to retrieve accumulation points of low intensities instead of the absolute minimum during the considered period of time. This approach has at least three advantages: First the algorithm is more robust against errors in level-1 data, especially if long periods of time are considered, because the result is not determined by one measurement alone. Moreover, the accumulation point method can take the seasonal variation of the albedo during the considered period of time into account, if there exist more than one measurement corresponding to cloud free pixels: the average of the intensities for cloud free scenarios is a better choice than the absolute minimum in this case. The third advantage is, that the minimum reflectance retrieved by an accumulation point method during long periods can be used as a pre-classification criterion for the analysis of short periods of time: Using long periods, we can identify all clouds which raise the intensity steeply to distinguish them from a variation of the surface albedo during the considered period of time. The assumed maximum variation of the surface albedo is pre-defined. The principle of the iterative fixpoint algorithm is shown in in the following diagram:



The algorithm is initialized by building up a set of daily global images containing the sum of the reflectances of PMD3, whereas all pixels with intensities clearly brighter than the Sahara are skipped. Each point of the image is now compared to the average image retrieved from the whole sequence. If the intensity of a measurement exceeds the sum of the average value and a predefined threshold, the measurement is interpreted as cloudy and skipped from the sequence. The result is an image sequence containing less clouds. This sequence is used as input to run the algorithm again. This is repeated until the image sequence does not change anymore. During stage 1, this algorithm is applied to the whole data set of SCIAMACHY measurements used for this release from January 2003 to November 2004. The result is used as input for the second stage and the average of this image sequence can be interpreted as first approximation of the lower threshold. During stage 2 and 3 the algorithm is applied to gradually smaller sets of SCIAMACHY data: stage 2 is applied to 4 different seasonal data selections; for stage 3 different SCIAMACHY datasets are selected for each day covering 37 days close to the day the thresholds are calculated for. After the third stage we obtain individual thresholds for each day, given by the average of the 37 days considered. Nevertheless, these maps received individually for each days by stage 3 contain several gaps corresponding to points, where no cloud free pixel is found during the 37 days. In this case, the algorithm has to use the average of the image sequences obtained from the earlier stages. Stage 4 can be used over deserts, but usually not for several other regions on earth. On the other hand, errors in the retrieved albedo lead to errors in cloud fraction, especially for deserts, because of the high albedo in the wavelength region used by HICRU. This makes higher precision over deserts useful. Nevertheless, HICRU uses the stage covering the shortest period of time that includes cloud free pixels.

## Thresholds for completely cloudy pixels (upper thresholds)

The upper threshold represents a completely cloudy pixel for a cloud with high albedo. Image sequence analysis is not necessary for the retrieval, because the thresholds do not depend on surface albedo. Therefore we retrieve the upper threshold dependent on solar zenith angle and the scan angle of the instrument. The algorithm works similar to the retrieval of the lower threshold, but is applied separately to 160 different data sets of PMD-measurements: Each data set covers PMD-measurements for a solar zenith angle bin of 3 degree (overall 20 bins). For each solar zenith angle bin, different values are calculated as a function of the scan angle (8 bins). The algorithm starts with all PMD-measurements from one of these data sets, whereas pixels definitely not representing completely cloudy pixels are skipped through a threshold method used for pre-classification. Afterwards, each measurement of the data set is compared with the average of all measurements. If a measurement underestimates the average of all measurements by more than predefined absolute and relative thresholds, it is removed from the data set. The result is a reduced list of PMD measurements, which are used to run the algorithm again. This is repeated until the list does not change anymore. The results show a significant dependency of the retrieved thresholds on both solar zenith angle and scan angle.

The choice of the algorithm's tuning parameters have to be selected carefully to obtain a smooth correlation between the upper threshold and the solar zenith angle without outliers due to events of single, bright measurements from clouds or ice surfaces. We use huge data sets (a whole year) in order to be mostly independent of climatological dependencies and robust to errors in the PMD data. A single measurement with very high intensity should hardly affect the result. The upper threshold represents completely cloudy pixels with a high, but not maximum or explicitly defined or retrieved albedo. Clouds with higher albedo than the "model cloud" represented by the upper threshold are interpreted as cloud fractions higher than 1 by HICRU.

Ice and snow covered surfaces can be brighter than clouds with high albedo. For the retrieval of the upper thresholds different, predefined regions usually covered by snow or ice are skipped.

# Auxiliary Data

No auxiliary data is used for HICRU. HICRU is completely based on SCIAMACHY data.

# Sensitivity, error analysis and validation

The errors depend on the surface, on the solar zenith angle and the quality of the level-1 spectra. The current release does not include an error for each measurement, this will be added for the next release. The errors are lower than 0.05, including the results over deserts which are often difficult for SCIAMACHY cloud algorithms. This is concluded in agreement with case studies, the analysis of the scattering for cloud free scenarios and the theoretically known error sources of the algorithm. For high cloud fractions the error analysis is difficult. Dependent on the conditions the algorithm can be more accurate. For small cloud fractions the error depends: on the availability of cloud free measurements for the retrieval of the lower threshold, the surface albedo and the seasonal variations of the surface albedo. The error is slightly higher for high scan angles, because -different from the HICRU algorithm for GOME- the retrieval of the lower thresholds does not take the dependency on the scan angle into account.

The algorithm does not work over snow and ice covered surfaces and in the case of sun glint. The results are overestimated in these cases.

The algorithm is compared to the FRESCO algorithm [4] using data from different seasons for validation. The correlation coefficient is higher over ocean than over land. The correlation coefficient is usually higher than 0.95, the differences are analyzed with respect to different parameters (e.g. solar zenith angle, surface type etc.). For selected data, the results from HICRU are also compared with Meteosat data and the validated HICRU GOME cloud algorithm. The results are unpublished, but presented on different conferences. Material can be found on our website [3]

# Recommendations for product validation

Effective cloud fraction cannot be compared directly to cloud data sets from surface observations or other satellite platforms, because most of these datasets do not provide an effective cloud fraction, but cloud parameters based on assumptions, which differ from HICRU and other cloud algorithms developed for the retrieval of tropospheric trace gases.

A validation should include an intercomparison of algorithms using different methods for the retrieval of effective cloud fraction (e.g. HICRU and FRESCO) covering different surface types (land, ocean, desert) and solar zenith angles. For correlation coefficients lower than 0.9 it can be assumed, that at least one of the algorithms may have a problem. The results should be compared to results from other satellite platforms too, but the limitations of these intercomparisions have to be recognized.

The HICRU algorithm is not too sensitive to calibration problems like an offset in absolute intensities, because the retrieval of effective cloud fraction is based on relative intensities. Nevertheless, because those problems are reported for SCIAMACHY, a successful implementation for GOME or a similar satellite platform with a good agreement to SCIAMACHY is also helpful for validation.

# References:

HICRU:

[1] Grzegorski, M., Frankenberg, C., Platt, U., Wenig, M., Fournier N., Stammes, P. and Wagner T.: Determination of cloud parameters from SCIAMACHY data for the correction of tropospheric trace gases, in: ESA publication SP-572 (CD-Rom), Proceedings of the ENVISAT and ERS symposium, 6-10 September 2004, Salzburg, Austria, 2004

[2] Grzegorski, M., Wenig, M., Platt, U., Stammes, P., Fournier, N., and Wagner, T.: The Heidelberg iterative cloud retrieval utilities and its application to GOME data, submitted to ACPD, 2005.

[3] http://satellite.iup.uni-heidelberg.de Choose "Data products" and "HICRU cloud data" on the website to find additional informations and download files.

Other references:

[4] Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W., and de Haan,J. F.: A fast method for retrieval of cloud parameters using oxygen A band measurements from the global ozone monitoring experiment, J. Geophys. Res., 106D, 3475–3490, 2001.

[5] Fournier, N., Stammes, P., de Graaf, M., van der A, R., Piters, A., Grzegorski, M.., and Kokhanovsky, A.: Improving cloud information over deserts from SCIAMACHY O2 A-band, submitted to Atmos. Chem. Phys., ACP 6, 163-173, 2006.