Cloud detection for the Advanced Infrared Radiometer Sounder (Part II)
- Proposals of new schemes and a study with the real data -

March 17, 2003

Yoshiaki Takeuchi
Japan Meteorological Agency
EUMETSAT NWP SAF Visiting Scientist to Met Office, UK

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 25 November 1998, between EUMETSAT and the Met Office, UK, by one or more partners within the NWP SAF. The partners in the NWP SAF are the Met Office, ECMWF, KNMI and Météo France.

Change record			
Version	Date	Author / changed by	Remarks
1.0	17.3.03	Y.Takeuchi	Version for distribution

Cloud detection for the Advanced Infrared Radiometer Sounder (Part II)
- Proposals of new schemes and a study with the real data -

March 17, 2003

Yoshiaki Takeuchi
Japan Meteorological Agency
EUMETSAT NWP SAF Visiting Scientist to Met Office, UK

Abstract

This report describes new cloud detection schemes named PCA scheme and Optional PCA scheme based on a principal component analysis for AIRS/AMSU-A onboard AQUA satellite and compare these schemes with a cloud detection scheme named Var scheme based on Bayesian theory and other schemes.

AIRS Focus day data are used to compare the performance of cloud detection and AIRS VIS/NIR high resolution images are used for the validation of these schemes.

Lower sensitivity of the multi-channel Var scheme to cloud in mid- and low-latitudes than a single channel Var scheme is solved by introducing the PCA scheme. The optional PCA scheme gives more symmetric statistics of observation minus background radiances by considering probability distributions of cloudy profiles.

Contents

1. Introduction

1.1 Importance of AIRS cloud detection study

Cloud detection is essential to retrieve accurate atmospheric parameters such as temperature and water vapor from satellites for Numerical Weather Prediction without considering cloud parameters such as cloud liquid water and cloud ice water in the retrieval because cloud affects observed radiance at the top of atmosphere through strong absorption, emission, and reflection particularly in the infrared spectral region.

The Atmospheric Infrared Sounder (AIRS) is an instrument onboard NASA's Aqua earth observation satellite launched on 4 May 2002. The AIRS has 2378 spectral channels at resolutions, $\lambda/\Delta\lambda$, ranging from 1086 to 1570 and among other things measures air temperature, humidity, cloud and surface temperature.

The Met Office has ingested near real-time AIRS brightness temperature (BT) data, named AIBT BUFR data, at 324 selected channels distributed by National Environmental Satellite, Data, and Information Service (NESDIS) and plans to use the data for meteorological applications. The distributed AIRS channels and their central wave numbers are listed in Table 1[1]. An advantage of AIRS over existing instruments in view of cloud detection is availability of many window channels free from gas line absorption with its very high spectral resolution. AIRS data in detected clear areas are used for bias correction of forward model and also for intercalibration among imagers with broad spectral response function onboard Geostationary meteorological satellites, i.e., METEOSAT, MSG, GOES, and MTSAT, because any sensor data with arbitrary response function can be constructed from the high spectral resolution AIRS data.

The AMSU-A microwave multichannel radiometer on Aqua is also used in this study. The AMSU-A has 15 channels measuring radiation in the frequency span of 23 GHz to 90GHz. The spatial resolution at nadir is 40.5km. Three channels (24GHz, 31GHz, and 89GHz) are predominantly used for atmospheric water vapor sounding. The remaining twelve channels (between 50GHz and 60GHz) are used for atmospheric temperature sounding.

1.2 Result from the simulation study

In Part I of this report (Takeuchi, 2002), a simulation study to obtain an optimum set of cloud detection channels and the thresholds for AIRS was performed with a sampled profile database based on the ECMWF 40-year re-analysis. The RTTOV-7 code and IASI-1DVar code is used to simulate AIRS, AMSU-A, and AMSU-B brightness temperatures for cloudy profiles and to calculate the cloud cost as a measure of cloudiness based on a Bayesian cloud detection scheme (English et al., 1999) named Var scheme. Sensitivity of the brightness temperatures to cloud and water vapour are calculated and five candidates for channel sets for cloud detection were obtained. Thresholds for each channel set is determined and cloud detection performance is verified with the true total cloud liquid and ice water.

The conclusions of the study were,

1) Cloud cost with single window channel tends to overlook ice cloud and lower-level warm cloud.
2) The 3.8 micron window single channel shows worse performance than the 10.5 micron single channel.
3) Combined use of long wave infrared window channels (10.5 micron and 8.9 micron) and short wave infrared window channels (3.8 micron) is essential to ice cloud detection.
4) Combined use of infrared and microwave channels is essential to low-level water cloud detection.
5) Very little improvement is obtained by using 7 AIRS channels cloud cost over 2 AIRS channels cloud cost.
6) If more channels are used in the Var scheme, smaller cloud cost is calculated and the cloud detection performance is degraded. The degradation is apparent at mid- and low-latitude.

A question raised in the prior report is ‘Why the Var cloud detection scheme with multi-channels results in worse performance than that with single channel at mid- and low-latitude although it does good performance at high latitude?’. In this study, we focus on the Var scheme problem, and investigate it theoretically and observationally with real AIRS data.

1.3 Treatment of undetected cloud

[1] NESDIS has included the visible information co-located to the AIRS footprints in the AIRS BUFR data (AIBT BUFR data) distributed to the U.K. Metoffice since 22 January 2003. See section 8.

A common problem in cloud detection with real data is the treatment of undetected cloud. The horizontal resolution of AIRS infrared channels at nadir is 13.5km, therefore, most of these data are affected by undetected small fraction of cloud or very thin cloud due to its low resolution. As a result, the probability of observation minus background brightness temperature is not symmetric.

For utilization for NWP assimilation, symmetric observation minus background brightness temperature statistics is desirable. If not, the analysis temperature or moisture fields are biased due to contaminated cloud. To sort out this problem, cloud probability distribution of cloud should be considered.

1.4 Real AIRS/AMSU-A data

The other important object of this report is to validate the performance of our cloud detection schemes and a few other schemes such as Mitch's scheme (Goldberg and Zhou, 2002) and ECMWF scheme (Watts et al., 2002) with real AIRS data. It is expected to find any problems on the real data and/or forward model through the validation. A sample full day data set called Focus day data provided by NASA/GSFC is used as a real AIRS data in this study.

Fig. 1 shows the general flow of this study with real AIRS/AMSU-A data. First step is a build-up the existing and proposed schemes into an IASI_1D-Var retrieval scheme. Second step is to perform cloud detection with AIRS/AMSU-A real data set. Third step is selection of validation case and preparing AIRS IR/NIR false color images. Finally the best choice of scheme are proposed as a Day-2 scheme from the validation results.

1.5 Motivation

In summary, the motivation of this report is to propose new schemes which give symmetric O-B[2] statistics for clear area and validate these schemes with AIRS VIS/NIR high resolution data. First, this report makes clear the meaning of the Var scheme proposed by English et al. (1999) and show a problem when it is applied to many channels. Then, to sort out the problem, a cloud detection scheme based on a principal components analysis is introduced and a new scheme in which the probability distribution of cloudy areas is considered is also proposed. In this report, we use the O-B statistics of derived clear profiles as a key property to validate the scheme.

1.6 Scope of this report

First, cloud detection per FOV, it means all channels are regarded as clear or not, is treated, so cloud detection per channel is out of scope in this study. Second, a scheme uses multi-FOV information is out of scope though it is very efficient in some cases. We focus on the utilization of the multi-spectral channels information given by high spectral resolution instruments such as AIRS and IASI. Third, cloud detection over ocean is only discussed because the foreward model is more accurate over ocean than over land. Forward model used here is not applicable under the solar illuminated condition, and so, short wave infrared channels are only used in the nighttime.

Results of this cloud detection are applicable to bias tuning and in 1D-Var retrievals and/or 3D-Var/4D-Var direct assimilation systems. Sensors and channels used are infrared channels of AIRS and microwave channels of AMSU-A. The AMSR, the other microwave instrument onboard Aqua, is not used since the FOV is difficult to be matched up with the AIRS FOV due to the difference of the scanning geometries. The relation of Fields of View of AIRS/AMSU-A is that an AMSU-A FOV encompasses 9 AIRS FOVs (arranged in a 3 x 3 matrix), and so, AMSU-A data is easier to co-locate.

1.7 Structure of report

In section 2, we make clear the reason of the problem theoretically and verify it with a simulation. Thereafter, a new scheme named PCA scheme is introduced in section 3. Another new scheme considering cloud probability disttibution named Optional PCA scheme is also described in section 3. In section 4, the details of the real data, background data, and validation data are described. The parameters used in each

[2] The difference between observed brightness temperature and background brightness temperature.

cloud detection scheme are also described in the section. In section 5 and section 6, the global performance and the case study are investigated, respectively.

2. Theoretical descriptions of Var scheme – To examine total cost of clear probabilities -

2.1 Formulation of Var scheme

The Var scheme based on Bayesian theory is applied as a Day 1 cloud detection scheme for operational near real-time AIRS data processing (Collard, 2002). It was first applied to HIRS and MSU data and is formulated as follows (English et al., 1999).

The cloud cost J_C that is a measure of cloudiness is calculated from the difference between the observed and background BT over the given channels as follows;

$$J_C = (\Delta \mathrm{y})^{\mathrm{T}} \left\{ \mathrm{H}(\mathrm{x}^{\mathrm{b}}) \mathrm{B} \mathrm{H}(\mathrm{x}^{\mathrm{b}})^{T} + \mathrm{R} \right\}^{-1} (\Delta \mathrm{y}) \tag{1}$$

where $\Delta \mathrm{y} = \mathrm{y}^{\mathrm{o}} - \mathrm{y}(\mathrm{x}^{\mathrm{b}})$, y^{o} is the observation and $\mathrm{y}(\mathrm{x}^{\mathrm{b}})$ is the estimated observation vector calculated from the background profile x^{b}, by a forward model, $H(\mathrm{x}^{\mathrm{b}})$. $\mathrm{H}(\mathrm{x}^{\mathrm{b}}) = \nabla_{\mathrm{x}} \mathrm{y}(\mathrm{x}^{\mathrm{b}})$ is the matrix containing the partial derivatives of the simulated observations with respect to the elements of x^{b}. B is the error covariance matrix of background NWP fields x^{b}, $\mathrm{R}=(\mathrm{E}+\mathrm{F})$ is the total observation error covariance matrix consisting of the instrumental error covariance matrix E and the forward model error covariance F.

Formula (1) can be rewritten as follows,

$$J_C = (\Delta \mathrm{y})^{\mathrm{T}} \mathrm{S}^{-1} (\Delta \mathrm{y}) \tag{2}$$

Matrix S represents the statistical structure of O-B values. Diagonal values of S are the variance of each channel and the off diagonal value means the covariance between any two channels. The representation of this formula is the same as that of well-known chi-square test. Chi-square test is a statistical test if summation of probability function with equal variance and mean. The mean of Jc is expected as N and the variance is expected as 2N where N is used channel numbers. In the Var scheme the clear case is declared if the Jc/N value is less than a given threshold. If used channels are independent of each other, the normalized cloud cost value Jc/N, hereafter we call it the 'normalized cost fuction', is independent of the number of the used channels. However, the BT difference of these channels generally have correlation. Therefore, the more channels we used, the lower the normalized cloud cost that is calculated.

2.2 Principal Component Analysis (PCA) of the cloud cost

Here, we calculate eigen values X and eigen vectors U of S as follows, then,

$$\begin{aligned} J_C &= (\Delta \mathrm{y})^{\mathrm{T}} \mathrm{S}^{-1} (\Delta \mathrm{y}) \\ &= (\Delta \mathrm{y})^{T} \mathrm{U} \mathrm{X}^{-1} \mathrm{U}^{-1} (\Delta \mathrm{y}) \\ &= \sum_{i=1}^{N} (\Delta \mathrm{y}_i')^2 \end{aligned} \tag{3}$$

Where $\Delta y'_i \left(= \sum_{j=1}^{N} \left(\Delta y_j\right)^T U_{ji} \Big/ \sqrt{X_{ii}} \right)$ is the i-th PCA components of Δy. Each PCA components are normalized by each eigen value and are orthogonal to each other. This formula implies that the Var scheme uses simple average of square of all PCA components and treats each PCA component as equal. In other wards, the Var scheme assumes each PCA component has same performance for cloud detection. Is it a valid assumption?

2.3 Validation of the Var scheme with a simulation dataset

2.3.1 Description of the simulation

To confirm the assumptions in the Var scheme, a simulation study is performed. At first, 'clear' O-B values and 'cloudy' O-B values are simulated from the ECMWF profile dataset[3] by RTTOV7[4] forward model. The observed brightness temperature denoted y^o is calculated for each ECMWF profile and total measurement error, i.e., instrumental error plus forward model error, is added to the result. The background brightness temperature denoted y^b is calculated from each ECMWF profile including background error.

In the simulation study, the background data are given by the ECMWF sampled profile database with background perturbations. The background perturbations are consistent with the error covariances expressed in the background error covariance matrix, B. As described in Collard and Healy (2002), this is done by decomposing the B-matrix,

$$B = X \Lambda X^T \tag{4}$$

And then calculating true-minus-background perturbation through

---

[3] The ECMWF sampled profile database (Chevallier, 2001) consists of 13495 atmospheric profiles sampled, i.e. geographically, temporally, and weather conditions, from the ECMWF 40-year re-analysis (ERA-40) data. The ECMWF sampled profile database also includes cloud properties, i.e. cloud liquid water, and cloud ice water, cloud cover, at ECMWF 60 model levels. The dataset characterises a regular distribution of physically consistent atmospheric temperature, water vapour and ozone profiles. Since these profiles are equally sampled in temperature, humidity, and ozone, the dataset is very useful, for example, in making a kind of regression coefficient for some retrievals and in this study. Atmospheric properties such as temperature, water vapour mixing ratio, ozone-mixing ratio are given at 60 model levels. This data set is used as an input to calculate cloudy Brightness Temperatures (BTs) of AIRS, AMSU-A, and HSB by RTTOV-7 as described in section 2.1. To match the interface, temperature, water vapour mixing ratio, ozone mixing ratio is interpolated onto 43 pressure level pre-determined in RTTOV-7.

[4] The Radiative Transfer for Tiros Operational Vertical Sounder fast radiative transfer model (RTTOV-7: Saunders et. al, 2002) is used to simulate the brightness temperature (BT) at each channel for cloudy atmosphere profiles in the ECMWF sampled database. Radiative transfer calculation is carried out by using temperature, water vapour and ozone volume mixing ratio profiles at 43 pressure levels and surface properties such as surface temperature, surface water vapour, and skin temperature and surface emissivity as input parameters.

The standard RTTOV model is not suitable for the simulation study because in this model clouds are assumed to be at one level, have unit emissivity and a top at a fixed cloud top pressure with a fractional coverage for each input profile. Therefore, we used RTTOVCLD routine which takes a profile of temperature, cloud cover, cloud liquid water and cloud ice water on user defined model pressure levels and computes infrared and/or microwave cloudy radiances for multilevel and multiphase cloud fields. In the RTTOVCLD code, cloud is assumed to be random-overlapped.

The simulation is carried out for 324 sampled AIRS channels, the Advanced Microwave Sounding Unit –A (AMSU-A) 15 channels. RTTOV-7 coefficients for NOAA-16 AMSU-A are used instead of AMSU on Aqua.

In this report, the satellite zenith angle is fixed at 0.0 deg (i.e. nadir view) assuming the channel selection and threshold value are insensitive to satellite zenith angle. Surface emissivity of IR channels is given by ISEM (Sherlock, 1999) for sea. Surface emissivity of MW channels is given by the FASTEM-1 (English and Hewison, 1998) for sea.

$$x^T - x^b = \sum_{i=1}^{N} a_i \sqrt{\lambda_i}\, x_i \tag{5}$$

Where the N eigenvalues λ_i, which are the diagonal values of Λ, are the associated eigenvectors (and columns of X). The a_i is from a set of normally distributed random numbers with unit variance and zero average. The background error covariance matrix used is the same as that used in the IASI_1DVAR code.

The observation data are given by simulating BTs based on the ECMWF sampled profile database in the prior step and then adding random noise that is consistent with the R matrix being assumed for AIRS, AMSU-A, and AMSU-B.

Total measurement error and background error used are the same as in the cloud cost calculation in the Var scheme. The 'clear' case here is defined as both of total cloud ice and total cloud liquid water content, those are calculated from cloud ice profile and liquit water profile the ECMWF profile dataset, are exactly zero while the 'clear' case in Takeuchi (2002) was defined as the case with total cloud liquid water less than 1.0 (g/m^2) and with total cloud ice water less than 0.1 (g/m^2). 6813 clear and cloudy profiles over the ocean are extracted from the ECMWF profile database and used for the calculation.

Second, the clear error covariance matrix for a selected channel set is constructed from clear O-B brightness temperature statistics and obtain the eigen values and eigen vectors as follows,

$$\begin{aligned} S_{CLR} &= \overline{(\Delta y_{CLR})(\Delta y_{CLR})^T} \\ &= U_{CLR} X_{CLR} U_{CLR}^{-1} \end{aligned} \tag{6}$$

where over bar means ensemble mean. This matrix represents the statistical structure of brightness temperature error consisting of total measurement error and background error. Channel sets used are mix02 and sound02 listed in Table 2. Channels used in these channel sets are selected based on the results from Takeuchi (2002). The sound02 channel set is a sub-set of mix02 without two short infrared channels, i.e., ch.2328 and ch.2333.

Third, PCA components associated with clear profile statistics are calculated for each clear and cloudy profile, respectively.

$$\Delta y_{CLR}' = (\Delta y_{CLR})^T\, U_{CLR} \sqrt{X_{CLR}} \tag{7}$$

$$\Delta y_{CLD}' = (\Delta y_{CLD})^T\, U_{CLR} \sqrt{X_{CLR}} \tag{8}$$

The result for mix02 channel set and sound02 channel set are shown in Table 3.1 and Table 3.2, respectively. The eigenvalues of clear error covariance matrix for mix02 in Table 3.1 a) shows that first four components are dominant and others are very small. The lower three components have variances almost the same as the total measurement noise, i.e. 0.04 K^2, it means these components have no signals from background profiles.

2.3.2 Results of the simulation

Table 3.1 b) shows mean and standard deviation of cloudy O-B brightness temperature associated with each clear PCA component. The large mean and standard deviation value means that cloudy profiles can easily devided from clear profiles. For the case of mix02 channelset, the upper nine PCA components are useful for cloud detection which have mean value larger than two and standard deviation value more than 2.7. The mean value is usually larger than the associated standard deviation value, which means the cloudy probability distribution and clear probability distribution is separated and can be divided by a threshold with ease.

On the other hand, the lower three PCA components are not useful for the cloud detection which have mean values less than three and standard deviation values less than two. The fourth PCA component has an exceptional characteristic, i.e. small mean value and large standard deviation, from the other useful PCA components. This component corresponds to brightness temperature differences between longwave infrared channels and short wave infrared channels, then, the component has opposite sensitivity by cloud particle phase, i.e. water or ice, as shown below.

Fig.2-5 are the scatter plots of each profile for 'clear' and 'cloudy' cases on two selected PCA components. Red plots are for clear case, green plots are for cloudy and total liquid water greater than total ice water case, and blue plots are for cloudy and total ice water greater than total water cloud case. The distribution of the plots looks like a comet with red coloured round core and blue and green coloured two long or short tails with different directions in some cases.

The red colored clear case plots distribute spherically around the origin. It means that the 'clear' PCA component for 'clear' O-B value is normalized its variance and not correlated each other. Even if a bit of cloud, for example the cloud liquid water less than 1.0g/m^2 or cloud ice water less than 0.1 g/m^2, is contarminated in the FOV, the result of the simulated O-B distribution departs from the pure clear case.

The most prominent feature to be noted in Fig.2-5 is the distribution of PCA components of cloudy O-B is much different by each PCA component as also shown in Table 3.1 b). For example, the first component has a cloudy distribution depart well from clear distribution. On the otherhand the twelveth component has a cloudy distribution which is almost embedded in clear distribution and we cannot divide clear FOV and cloudy FOV with this component.

In detail, for the case of mix02 for ocean, 1) high sensitive components to cloud are PCA02,06,07,09, 2) moderate sensitive components to cloud are PCA05,01,03,08,04, and 3) sensitivities of PCA10-12 are very low. In conclusion, PCA01-PCA09 are to be significant for cloud detection over ocean.

For sound02 channel set (Table 3.2), the clear cases are characterized by upper four PCA components and the last two components have almost no signal from background information. The PCA07 and PCA08 are almost no use for cloud detection because the mean and standard deviation of O-B value for cloudy case in the PCA space associated with clear case are very small. Like the case of mix02 channel set, the upper six components are useful for cloud detection with large mean and large standard deviation. The fourth component are similar to the fourth component of mix02 case, that is, the components is sensitive to cloud particle phase. The characteristic to be concerned is the tenth component, which has almost no signal with clear background information but the mean and standard deviation of cloudy case is large, that means this component is somewhat useful for cloud detection. As shown in the scatter plot in Fig.5 some cases of the cloudy distributions are discernible from clear distribution (upward tail) though the other cases are embedded in the clear probability distribution (dirty core). So we should pay attention to even lesser components which have some sensitivity to cloud. These lower components are to be considered not as the "unusable component" but as the "components free from background signal or perturbation". When such a component has some sensitivity to cloud, the component is useful to detect cloud.

In detail, for the case of sound02 for ocean, 1) high sensitive components to cloud are PCA01,02,06, 2) moderate sensitive components to cloud are PCA03,04,05, and 3) the sensitivities of PCA07-10 are very low. In conclusion, PCA01-PCA06 are useful over ocean.

As described later, we can regard the cloud test in Mitch's scheme as a scheme based on the similar concept. A cloud detection scheme with virtual channels (ref.) is also regarded as a similar scheme of the same family. Then we would think a PCA based scheme is a comprehensive and extensive scheme. Here, we have two choices to implement the PCA scheme for cloud detection. The first is to use the global probability statistics based on clear O-B statistics. The second is to use the local probability statistics dependent on the local profile. The first option cannot consider the difference of the vertical profile of temperature and water vapor which affects the background term in formula (1). Therefore we choose the second option so that the PCA calculation is carried out for each FOV.

We should note that the result is a simulated one and the effects of forward model bias and solar radiation. The components also tend to be degraded by forward model bias. It should be noted that the PCA components with the small eigenvalues tend to be sensitive to the spectral dependent model bias.

Here we have the answer to the question posed in Takeuchi (2002). Each PCA component has a different sensitivity to cloud, however, the Var scheme assumes all PCA components have the same performance as we pointed out in the section 2.2 so that uneffective components such as the PCA10-12 for the mix02 channel set are ineffective to detect cloud, therefore, the simple avarage of cloud cost tend to be small values. As a result, the cloud cost in Var scheme with multi-channels is ineffective to detect cloud. For this reason, a PCA scheme is introduced in section 3.1.

The second feature we should note in Fig.2-5 is that the distribution of cloud profiles on the PCA plot is heterogeneous and unsymmetric. This feature is important but not considered in the Var scheme and also in the PCA scheme introduced later. In other words, these schemes are to be called not as “cloud detection” but as “clear detection”. An Optional PCA scheme considering cloudy probability distribution is introduced in section 3.2.

3. New schemes based on Principal Component Analysis

3.1 PCA scheme – To examine clear probabilities for each PCA components -

As described in section 2.2, the Var scheme cannot treat PCA components of the cloud cost individually. Therefore, when many channels are used, the cloud detection performance of the Var scheme is degraded by averaging all partial cloud cost $J_{Ci}\left[=\left(\Delta \mathrm{y}_i'\right)^2\right]$ in Eq.(3), each of which is associated with the PCA component, even if some of the PCA components are more efficient to discriminate clear and cloudy areas. This is the reason why the Var scheme with many channels shows poor performance especially for low level cloud at low latitude to which most channels tend to have a similar sensitivity.

To solve this problem, each PCA component is examined individually in our PCA scheme. In general a couple of (higher and lower) thresholds for each PCA component can be given. Fig.6 illustrates the PCA scheme for a two-dimensional case. The red contours means cloud cost of clear FOV and green contours means cost of cloudy FOV. The first PCA component is in the vector (1,1) direction and the second PCA component is in the vector (1,-1). We can see that abilities to discern clear FOV and cloudy FOV are better with the first PCA component than with the second one. Because the Var scheme considers only the average of all partial cloud costs for clear, the scheme is not efficient in this case. The PCA scheme can apply the thresholds for the first PCA component and does not need averaging, the cloud detection efficiency is not degraded.

Though the thresholds can be assigned arbitrarily, the simplest way is to apply a constant value as all PCA components are normalised by the eigenvalues. This is equivalent to applying the thresholds parallel to the tangents of the contour with a cloud cost for a limited number of partial cloud costs. It should be noted that a similar result will be obtained if the single threshold is applied to all partial cloud costs. To use the partial cloud cost instead of the PCA component itself means that this simplest PCA scheme does not consider the asymmetricity due to the cloud probability distribution.

In the PCA scheme, the PCA component is calculated for each FOV based on Eq.(3) to consider the difference of background profiles from place to place. The background error covariance matrix and the total measurement covariance matrix used in the PCA scheme are the same as those used in the Var scheme.

An issue to be considered is the additional computer time for the eigenvalue and eigenvector calculation for each FOV. The quantitative estimate of the additional computer time is described in section 5.4. As a preliminary insight, the introduction of the eigen value calculation does not require much computer resource because the number of channels for cloud detection are not large, the order is ten to twelve, therefore, the computer time is negligible in the whole procedure including the iterative calculation of the temperature and water vapor retrievals.

For real AIRS/AMSU-A data, the procedure is as follows,

1) calculation of background brightness temperatures y^b of channels used by using RTTOV7 with background profile, i.e. atmospheric temperature, water vapor, and ozone profiles and surface properties such as surface temperature, surface wind speed. These input parameters are produced

based on a six hour forecast from operational NWP model at the Met Office. The K-matrix $H(x^b)$ is also calculated simultaneously with the same profile.

2) calculation of Δy by subtracting y^b from the observed brightness temperature y^o.
3) calculation of matrix S by using predetermined background error covariance matrix B and total observation error covariance matrix R and K-matrix $H(x^b)$ calculated in step 1)
4) calculation of eigen value and eigen vector of S.
5) all partial cloud costs J_{Ci} are calculated by using Eq.(3),
6) a FOV is declared as not clear if any principal component exceeds a given threshold. A value of 2.0 is used in this study noting that most of clear cases have all of PCA values less than 2.0 as shown in Fig.2 and Fig.4.

3.2 Optional PCA scheme – Comparison between clear probability and cloudy probability -

The Optional PCA scheme is an extension of the PCA scheme which considers the probability distribution of cloudy profiles in O-B space. The PCA scheme described in the previous section can detect more cloud in tropical region, however, the O-B statistics still have a weak asymmetry due to contamination of small fractions of cloud or very thin cloud in the clear FOV. As a result, the mean value of the O-B brightness temperature in the clear FOVs is generally biased to negative in the infrared and positive in the microwave spectrum region. To relax this deficit we introduce the probability distribution statistics of cloudy profiles in the O-B space in the Optional PCA scheme.

Fig.7 illustrates the Optional PCA scheme for a two-dimensional case. The red contours and green contours denote cloud cost based on clear case and the cost based on cloudy probability respectively, as Fig.6. Blue contour means the both costs have a same value. In the Optional PCA scheme, this contour is used as the boundary for the cloud detection. As shown in Fig.7, the optional PCA scheme, by cutting off the cloudy side and putting in the anti-cloudy side based on a statistical prior knowledge on cloudy probability distribution, can obtain rather unbiased O-B histogram.

The boundary of clear and cloudy is changed by clear probability distribution and cloudy probability distribution. A special case is shown in Fig.8, in which the covariance of cloudy probability is strongly correlated to each other. In this case, the second PCA has no information on clear/cloudy classification. We should note that this does not mean the second channel is no use. The performance with only the first channel is worse than that with two channels.

The procedure for real data are follows,

1) Prior to the processing the real time data, probability distribution statistics of cloud profiles are calcluated based on the previous day. The distribution is parameterized by the mean value and the variance of each cloudy PCA component. The cloudy PCA components are calculated from O-B of cloudy profiles.
2) Partial cloud costs based on clear statistics, i.e. total measurement error covariance, is calculated by the same way used in the PCA scheme, thereafter, the partial cloud costs are accumulated over the first six PCA components.
3) Cost based on cloudy statistics is calculated by accumulating the first six PCA components obtained.
4) To subtract cloud cost based on clear statistics from cost based on cloudy statistics and if the result is larger than the given threshold, 0.0 is used in this study, the FOV is declared as clear.

3.3 Brief descriptions of other schemes used for the validation

To compare with our cloud detection schemes, two major schemes, namely Mitch’s scheme and ECMWF scheme have also been implemented. The characteristics and theoretical background of these schemes are described in this section.

3.3.1 Mitch’s cloud detection scheme

Mitch’s cloud detection scheme (Goldberg and Zhou, 2002) uses AIRS ch.791, 843, 914, 1285, 1301, 2112, and 2226 and AMSU-A ch.4, 5, and 6. The scheme consists of three tests namely test1A, test1B, and testSST. The most important feature of this scheme is not to use a priori values except for sea surface temperature (SST).

The Test1A is a comparison between retrieved AIRS ch.2112 (2391.1cm-1, 4.18micron) brightness temperature and observed AIRS ch.2112 brightness temperature. The retrieved AIRS ch.2112 brightness temperature is estimated with AMSU-A ch.4, 5, and 6. The difference of scan angle dependency and the difference of solar irradiance dependency between the infrared channel and the microwave channels are also considered as belows,

AIRS_2112_Ret=18.653-0.169*AMSU_4+1.975*AMSU_5-0.865*AMSU_6
+4.529*(1.-COS(Scan))
+0.608*COS(SolZen) (10)
test1A = AIRS_2112_Ret - AIRS_2112 (11)

where Scan is scan angle and SolZen is solar zenith angle in radian at the FOV. The last term of Eq.(10) is included only in the daytime, i.e. SolZen is less than $\pi/2$.

In the Test1B short infrared AIRS ch.2226 (2532.0 cm-1, 3.95micron) brightness temperature is compared with long infrared AIRS ch.843 (937.92cm-1, 10.7 micron) brightness temperature.

test1B = AIRS_2226 - AIRS_843 (12)

TestSST is a comparison between retrieved SST temperature and a background SST value. Retrieved SST temperature is calculated by using AIRS ch.791 (918.76cm-1, 10.88micron), 914 (965.44cm-1, 10.36micron), 1285 (1228.2cm-1, 8.14micron), and 1301 (1236.6cm-1, 8.09micron). These channels are long wave infrared channels and can be used to estimate a strict SST by cancelling out a contribution from water vapour absorption to brightness temperature by the multi-channel regression.

SST_Ret = 8.28206-0.97957*AIRS_791 +0.60529*AIRS_914
+1.74444*AIRS_1285 -0.40379*AIRS_1301 (13)
testSST = SST(background) - SST_Ret (14)

All of the three tests are applied with given thresholds. In this study, the following thresholds are used regarding Goldberg and Zhou (2002), as follows,

test1A_thresh=2.0, test1B_thresh=5.0, testSST_thresh1=2.0, testSST_thresh2=4.0

A FOV satisfying the following all of these tests is declared as clear and otherwise declared as cloudy.

test1A < test1A_thresh
test1B < test1B_thresh
testSST > testSST_thresh1
testSST < testSST_thresh2

Mitch's scheme is an alternative candidate when we cannot use background profiles for any reason. The other advantage of the scheme over other schemes in this study is almost no need of computer memory and CPU time because this scheme does not need a forward calculation. However, from a theoretical point of view, we can point out a few weakness of this scheme.

First, the test1A and the test1B use the constant thresholds, then, it is possible that the scheme does not work well in an extreme case such as with a temperature inversion. In addition, the test1A and the test1B use short wave infrared channels which are affected by solar irradiance, therefore, it cannot be used in the daytime.

Second, AIRS ch.2112 and AMSU-A ch.5 used in the test1A have the peak of their weighting function at around 4 km in height, therefore, the brightness temperature sensitivity to cloud is not too large, therefore, the cloud signal cannot be detected efficiently especially for low level cloud although these channels are not influenced by ground surface temperature and/or emissivity variability though they are better for high cloud.

Third, the testSST assumes the difference between retrieved SST and background SST ranges between 2K to 4K. However, the difference depends on vertical atmospheric profile, so, in an extreme case such as a warm surge over cold ocean, the difference could be a negative value and the testSST gives an incorrect result.

3.3.2 ECMWF scheme

The ECMWF scheme is a scheme to give cloud unaffected channels, in other words, cloud detection per channel, developed by ECMWF (Watts, *et. al.*, 2002). This scheme makes use of forecast model clear radiance estimates.

The characteristics of the ECMWF scheme are easy to implement, almost no additional computer resources because the input variables required in this scheme are also used in the assimilation system.

The procedure of the original ECMWF scheme are as follows,

1) Radiation effect of black cloud defined as $\left|R_{CLD}(p_i) - R_{CLR}\right| / R_{CLR}$ is calculated for each channel with a background atmospheric profile, where R_{CLR} is the background radiance and $R_{CLD}(p_i)$ is the radiance calculated by assuming a black cloud located at each model pressure level p_i.
2) Cloud unaffected radiances for p-level is calculated for each channel, at the level the radiation effect has the same as a threshold, i.e. 0.01. This cloud unaffected p-level is obtained as a real value by interpolation from the model pressure levels p_i.
3) All AIRS channels are sorted with the cloud unaffected p-level for each band, namely, 15micron, 7micron, 4.5micron, and 4.2micron bands. This procedure is called 'channel pressure ranking', and the index of the sorted channel is called 'ranked channel index'.
4) The difference of the observed brightness temperature O and the background brightness temperature B is sorted by the 'ranked channel index', thereafter, a low-pass filter is applied to the O-B over 'ranked channel index'. About 10 channels are applied as the filter band width.
5) By thresholding the gradient of the filtered O-B value to the 'ranked channel index', clear channels are declared for each band. In the original ECMWF scheme, the threshold of 0.01 is used for the gradient check.

From a practical point of view, the ECMWF scheme has the following problems.

First, the scheme requires a forward calculation for all useable AIRS channels even when all the channels are not used in the following retrieval step, therefore, the scheme needs more computer time to perform the forward calculation than the other schemes which use the selected channels only. If only selected channels are used in this scheme, the low-pass filtering and the O-B gradient test is difficult to apply.

Second, the scheme is very sensitive to the accuracy of the background profile and the forward model, because the threshold of the gradient test is very strict, i.e. 0.01. This strict value is given based on the assumption that a BT difference less than 0.1K due to cloud contamination is acceptable for satellite radiance assimilation for NWP. As a result, most of 'true' clear profiles are possibly rejected when some biases are included in the background profile and/or forward model, and so, a strict bias correction of forward model is essential to apply this scheme successfully.

Third, the scheme does not use the background error covariance between the different spectral bands. As a result, a FOV with small O-B value near zero tends to be declared as a clear. As shown in Takeuchi (2002), the difference of sensitivity of brightness temperature to cloud among longwave IR band, shortwave IR band, and MW band is very useful to identify cloud. Though neglecting the background error covariance contributes to saving the calculation time, the useful information is overlooked in the scheme.
As a result, the ECMWF scheme tends to miss very thin cloud such as cirrus and homogeneous low level cloud the temperature of which is almost the same as the underlying surface.

Fourth, the physical meaning of thresholds of gradient of O-B value against 'ranked channel index' is not clear, and the optimization of the threshold is rather empirical.

These problems will be discussed in section 6.

The the program code of the original ECMWF scheme is not able to be provided because the scheme is not a kind of subroutine but embedded in the assimilation system, it been coded as a subroutine of the IASI_1Dvar standalone retrieval program. Basic performance was confirmed by the simulation data set with/without background and observation error.

3.4 Summary of the schemes

Table 4 summarizes the characteristics of the schemes used in this study. The Var scheme and the PCA scheme are the same except for the thresholding since these schemes use cloud cost associated with background error covariance as shown in Fig.6. Both schemes assume the cloudy probability is a constant everywhere in O-B space. On the other hand, the Optional PCA scheme considers cloudy probability distribution in the O-B space.

Mitch's scheme does not use any background properties except for sea surface temperature. However, the tests in the Mitch scheme utilize some relationships between channels in different spectral bands with the thresholds. For instance, the test1A uses the relationships between short wave IR channel and MW channels, the test1B the relationships between long wave IR channel and short wave IR channel, testSST the relationships between observed SST obtained by a regression and background SST. In addition, the thresholds reflect the background error covariance and measurement error independent of a forward model. In this sense, we might regard the scheme as a limited version of PCA scheme.

ECMWF scheme is also regarded as a special version of PCA scheme which uses only single spectral bands but all channels in the band. The sorting of channels and the low pass filter is equivalent to using a feature represented as a linear combination of the original channel. The gradient of the feature is also regarded as a feature. To perform cloud detection for a channel the twice features, i.e., low pass filterd O-B themselves and their gradients, of the channel width of low pass filter are used in the scheme.

4. Validation

4.1 Real dataset on a Focus day and the background data to perform cloud detection

4.1.1 AIRS IR and AMSU-A data

The sample full day data set of on-orbit Level-1B AIRS/AMSU/HSB radiances for 20 July 2002. AIRS/AMSU-A real data for 20 July 2002 are provided as a Focus day data by NASA/Goddard Flight Space Center for preliminary science evaluations.

On the Focus day data, the AIRS is in nominal science mode and the data quality is good. The preliminary validation by NASA indicates an absolute radiometric accuracy of better than 0.5K. The bias of channel 2333 (2617.5cm-1) is estimated as about -0.3K. The radiometric sensitivity is excellent. The noise equivalent temperatures are about 0.1K for long wave infrared window channels and about 0.3K for short wave infrared window channels.

The AMSU-A is also in normal science mode and the data quality is good. The calibration accuracy specification is on the order of 1K and less than 1K in orbit. The on-orbit noise equivalent temperatures are about 0.2K for channel 2 (23.8GHz) and 3 (31.4GHz) and about 0.1K for channel 15 (89.0GHz). AMSU-A channel 7 exhibits abnormal noise levels but the channel is not used in this study. The AMSU-A (channel 3 to channel 15) exhibits substantial scan asymmetry because no sidelobe correction has yet been applied. Off-nadir data (scan angle > 15-20deg) should be used with caution.

The Focus day data in this study include Level 1B brightness temperature for AIRS 324 selected channels and for simultaneous AMSU-A 15 channel data provided by NESDIS. Observation data are decoded from the BUFR data and formatted as the input data of IASI_1Dvar program. The observation data include latitude, longitude, elevation, surface type, satellite viewing angle, solar zenith angle, AIRS (324ch) and AMSU-A (15ch) observed brightness temperature at each FOV. These observation files contain all observations where |O-B| for channel 787 is less than 5K over open ocean. The one day data are assembled in six hourly, namely 21-03, 03-09, 09-15, 15-21UTC, and called 00, 06, 12, 18UTC data, respectively. The number of the FOVs in the data is 36438, which consists of daytime data of 17313 and nighttime data of 19125.

4.1.2 Background data

The background data for the focus day are produced by the Operational Processing System at the Met Office. The data include temperature, water vapor volume mixing ratio, and ozone volume mixing ratio at 43 pressure levels and surface temperature, surface humidity, skin temperature, surface pressure, and U and V components of winds at 10m height. These properties are extracted from the six hour forecast product of the NWP model at the UK MetOffice. The highest pressure level of the model is 10 hPa, so the temperature at upper seven levels, i.e. 6.95, 4.407, 2.611, 1.42, 0.69, 0.29, and 0.10 hPa, is given by a regression of AIRS stratospheric sounding channels. Water vapor volume mixing ratio higher than 122.04 hPa is given as a constant for each FOV. For details see Collard (2002).

4.2 Validation data sets

4.2.1 PCA image of AIRS and AMSU-A cloud detection channels for cloudy case

To interpret the difference between cloud detection results, the Primary Component Analysis of O-B values is conducted and the PCA images are produced. By this analysis, we can estimate how much information content the original O-B data includes and we can also guess the reasons of the variance of each PCA components from its geographical pattern and amplitude.

4.2.2 Geostationary satellite images

In the early stage of this study, Geostationary satellite visible and infrared images were used to browse the largescale cloud distribution. GOES-8, GOES-10, METEOSAT-7, METEOSAT-5, and GMS-5 images were all used to select the case study regions.

4.2.3 AIRS VIS/NIR false composite image and enhanced image

4.2.3.1 General descriptions of AIRS VIS/NIR photometer

As a primary validation data set, real AIRS Visible/Near-IR (VIS/NIR) false composition browse images provided by NASA/GSFC DAAC were used. On the focus day, the AIRS Visible/Near-IR (VIS/NIR) photometer was in normal science mode.

The photometer contains four spectral bands, the spatial resolution which at nadir is 2.3km. The primary function of the AIRS VIS/NIR channels is to provide diagnostic support to the infrared retrievals: setting flags that warn of the presence of low-cloud or high variable surface features within the infrared field-of-view. Each AIRS FOV encompasses 72 Vis/NIR pixels (arranged in a 9 (along track direction) x 8 (cross track direction) rectangular array).

The channel 1 (0.40 to 0.44 micron) is designed to be most sensitive to aerosols. Channels 2 (0.58 to 0.68 micron) and 3 (0.71 to 0.92 micron) approximate the response of AVHRR channels 1 and 2, respectively, and are particularly useful for surface studies (AVHRR is an imaging instrument currently carried by NOAA polar orbiting satellites). Channel 4, not used in this study, has a broadband response (0.49 to 0.94 micron) for energy balance studies.

The AIRS VIS/NIR photometer is characterised by high spatial resolution and the different spectral bands in visible and near infrared region so that we can easily validate low water cloud and also high cloud in the daytime.

4.2.3.2 AIRS VIS/NIR false composite image

AIRS VIS/NIR false image is a composite of blue for ch.1, green for ch.2, and red for ch.3. In this false image, black or dark blue area corresponds to open sea without 'thick' cloud. Bright white color corresponds to thick cloud and light grey color corresponds to thin cloud. The color of thin cloud has rather whitish veils. Thin middle or upper cloud can be also identified by blurring of underlying cloud edge. Shadow of upper cloud onto underlying cloud or surface is also useful to detect cloud.

Sun glint area shows a glowing appearance over the ocean. Aerosol such as Saharan dust, Yellow sand, and smoke of forest fire look like yellow veil over ocean and land. It is worth noting that the scan edges are light bluish over open sea due to Rayleigh scattering by air.

4.2.3.3 AIRS VIS/NIR enhanced image

The original AIRS VIS/NIR false composite image is efficient to identify low cloud but not good at thin cloud detection because such cloud has a very small optical thickness even in visible spectrum, namely the cloud is transparent. Therefore, we studied how to visualize such very thin clouds with the AIRS VIS/NIR composite image and found that enhancing the ch.3 shows the pure clear area as black color and the area covered very thin cloud as red color over ocean.

In general, cloud has flat spectral reflectance over VIS/NIR region and surface reflectance and Rayleigh scattering has significant spectral dependency over the region, i.e. NIR reflectance is smaller than VIS reflectance except for cloud, therefore, the enhancement of ch.3 makes the very thin cloud visible as red colored area. It should be noted that clear area in sun glint also shows light red appearance even without cloud.

The AIRS VIS/NIR enhanced images used in this study are produced by enhancing the lower part (0-20) of red color table (the original full range is 0-255) for original AIRS VIS/NIR false composite image to the full range of 0-255.

As shown in the case study in section 6, the red colored area over ocean can be regarded as very thin cloud, haze, or aerosol with large particle size. Though the height of those particles cannot be assigned by VIS/NIR images, the associated cloud type (cumulus, cumulonimbus, or cirrus), blurring the underlying cloud or not, the sharpness of the edge and the relationships with coast line give us some information on the height and composition. An area with yellow color in the original image and with orange color in the enhanced image can be regarded as an aerosol layer with small particles such as desert dusts and forest fire smoke.

4.2.4 Mitch’s scheme result

The details of Mitch’s scheme were described in section 3.3.1. Mitch’s scheme does not use background profile information except for sea surface temperature, therefore, we can show the advantages of using background profile information in the Var, PCA, and Optional PCA schemes by comparing these results with Mitch's scheme result.

4.2.5 ECMWF scheme result

ECMWF scheme’s feature is to treat the O-B values in different spectral bands independently. Therefore we can confirm the advantages of the multi-band cloud detection schemes, i.e. the Var, PCA, and Optional PCA schemes, against a single-band cloud detection scheme by comparing them with ECMWF scheme. To ensure this, the ECMWF scheme in this study is modified by relaxing the thresholds for cloud detection because the original ECMWF scheme gives only small clear fraction which makes the characteristics of single band cloud detection difficult to see. In addition, another modification is necessary because the original ECMWF scheme is very sensitive to the accuracy of the forward calculation; the scheme is available only by using an accurate Instrument Spectral Response Function (ISRF). The details of the modification of ECMWF scheme are described in the next sectoin. In this study, a FOV is regarded as clear if AIRS ch.914, which is a window spectral region, is declared as clear in the ECMWF scheme.

4.3 Implementation of cloud detection schemes and the parameters used

All Var, PCA, Optional PCA, Mitch, and ECMWF schemes are implemented into the IASI_1Dvar system version 2.3, which is a stand-alone retrieval system for sounders such as ATOVS, AIRS, and IASI (Collard, 2002) and the input and output interface are common for all schemes. The IASI_1DVAR code is a program developed at the Met Office as part of the Numerical Weather Prediction Satelllite Application Facility (NWP SAF) to retrieve atmospheric properties such as temperature and humidity by using a 1DVar scheme. The code was originally developed for the IASI instrument but it can be used with many different sounding instruments such as AIRS and AMSU. By changing the channel selection for cloud detection, cloud cost for any channel combination can be calculated. One of the great benefits of the IASI_1DVAR code is the capability to treat multi-sensor data such as AIRS and AMSU simultaneously.

The input data of IASI_1DVAR code are background profiles at 43 pressure levels, observation BTs of the instruments, the background error covariance matrix, and observation and forward model error covariance matrices. The radiative transfer model for AIRS and AMSU in IASI_1DVAR is based on RTTOV7 except that it does not include the cloudy radiation simulation function.

Equivalence between RTTOV7 code used for cloudy BT simulation and RTTOV7 code in IASI_1DVAR code are verified using clear profiles. To process a large profile dataset, the code is modified so that the background profiles are given one by one, rather than all together. In addition, a kind of profile modification as a pre-processing of the IASI_1DVar is removed from the code to avoid producing an unexpected cloud cost because the modification is implemented to get better retrieval results.

4.3.1 Mitch's scheme

In this validation, observed SST, an input parameter of Mitch's scheme, is given from the skin temperature in the background data. The same coefficients and thresholds as the original scheme are used without any change.

4.3.2 ECMWF scheme

The equivalent forward model based on the RTTOV-7 in the IASI_1Dvar system is used to calculate background brightness temperature for ECMWF, Var, PCA, and Optional PCA schemes. The '20 July' version of the ISRF is used. This version is much better than the pre-launch ISRF.

As noted in section 4.2.5, the ECMWF scheme is modified in this study. The Blackman low-pass filter is used for smoothing O-B value over 'ranked channel index' in the modified ECMWF scheme; the weight of the filter *w* is,

$$w = 0.42 + 0.5\cos\left(\frac{2\pi i}{N-1}\right) + 0.08\cos\left(\frac{4\pi i}{N-1}\right) \qquad \left(i = -\frac{N-1}{2}, \frac{N-1}{2}\right) \tag{15}$$

where, N is the filter band width and 11 is used in this study. At first, the gradient threshold of 0.01 on gradient of O-B value and the 'gross check' threshold of 2.0 on absolute O-B value are applied to find a cloud unaffected p-level. These thresholds are the same as the original ones.

The ECMWF scheme is applied by keeping very tight gradient thresholds experimentally to Focus day data. Unfortunately, the original thresholds does not work well for the real AIRS data and almost all FOVs are declared as cloudy. 'Virtual' cloud is detected in the stratosphere due to large O-B biases at stratospheric channels. Though the best way to solve the problem is introducing a correct bias correction it is not available here, so, to avoid the problem, 1) cloud check for stratospheric channels, the cloud unaffected p-level of which is larger than 17, is skipped, and 2) the threshold of gradient value of O-B value is set as 0.5 instead of the original value of 0.01.

Fig.9 shows some examples of the calculation with the modified ECMWF scheme. All channels of 9.6micron ozone band and 4.5micron band are semi-transparent, i.e. the weighting function is spread into ground surface, therefore, the scheme can not applied. 7 micron water vapor band has many sounding channels and the scheme can applied, however, channels sensitive to upper troposphere show large positive O-B bias mainly due to poor background water vapor profile as shown in Fig.9 a). As a result, the determined cloud unaffected pressure level is much higher than of the other bands.

The upper sounding channels in 4.2 micron temperature band also shows large positive O-B bias mainly due to non LTE effect as shown in the daytime as shown in Fig.9 c) and d). And the 3.8 micron window channels are affected by solar irradiance. These biases makes cloud detection absolutely difficult and it is concluded that the implementation of ECMWF scheme relies on correct forward calculation and/or bias correction.

15 micron temperature sounding band has many channels sensitive to stratosphere and the bias is not small and the low pass filtered O-B values are flat over 'ranked channel index'. And the gradient of the O-B values to 'ranked channel index' is larger in the cloudy case as shown in Fig. a) and b) and almost zero in the clear case as shown in Fig.9 c) and d). Therefore, this study uses only 15 micron long wave infrared band in ECMWF scheme in cloud determination.

4.3.3 Channels used in Var, PCA and Optional PCA schemes

Two cloud detection channel sets, i.e. mix02 and sound02, are commonly used in the Var, PCA, and Optional PCA schemes. These channels are the same as used in the simulation study in section 2.3. The sound02 channel set is a subset of mix02 without two shortwave infrared IR channels, i.e. ch.2328 and ch 2333. In this section, the reasons of the selection of the channel set are described.

In Part I of this study, the 'mix+amsu' channel set is nominated as an optimum channel set. It consists of AIRS ch.787, ch.843, ch.914, ch.1221, ch.1237, ch.2328, ch.2333, and AMSU-A ch.2, ch.3, and ch.15. All of these channels are window channels with large sensitivity to cloud. These channels have three significant bands sensitive to cloud around 10.5 microns (ch.787, ch.843, and ch.914), 8.9 microns (ch.1221 and ch.1237), and 3.8 microns (ch.2328 and ch.2333). These three window bands are suitable for cloud detection. In particular, channels in the 10.5 micron bands have the highest sensitivity to cloud. For cloud detection, the clouds to be concerned are thin ice clouds, which are semi-transparent even to infrared channels, and lower level clouds, the top temperature of which are close to surface temperature. Sufficiently thick cloud or much colder cloud than surface temperature are easy to detected with any simple scheme.

Water vapor sounding channels and ozone channels are rejected in the channel selection because the accuracy of the water vapor profile and ozone profile forecasted by a NWP model is insufficient to give a reliable background brightness temperature. Including these channels causes a noisy cloud detection result depending on water vapor and ozone variability. The above three window bands have rather smaller sensitivities to water vapor. In particular, channels in the 3.8 micron bands are almost free from water vapour perturbation because the 3.8 micron (2610 cm^{-1}) channels are located around the minimum of the water vapour continuum (Kneizys et al.,1980). The AIRS channel (2616cm-1) used in the Evan Fischbeins' clear sky detection is almost the same as this channel. A key point is that the 10.5 micron band and 3.8 micron band have different sensitivity to cloud and water vapor, so the combination of these bands give better results than the single band in cloud detection.

Three microwave window channels are included in the 'mix+amsu' channel set. Though the sensitivity of cloud in the microwave spectral region is smaller than at infrared wavelengths, brightness temperatures of microwave channels give independent information about cloud because the sea surface emissivity is much less than unity and cloud emissivity is lower than that in the infrared region. Therefore, microwave channels can detect low-level cloud even if the temperature of cloud surface and ground surface are almost the same. In the simulation study, it was shown that many rather thick liquid water cloud can be detected by adding AMSU-A ch.2, ch.3 and ch.15.

<u>Long wave infrared sounding channels</u>

Based on the results above, the additional possibilities are studied on temperature sounding infrared channels based on O-B difference of real AIRS data. As well known, long wave sounding channels around 10-15 micron and short wave sounding channels around 4.2 micron are used for retrievals of temperature profiles. These sounding channels have an advantage because of the insensitivity to surface property variability. Fig.10-1 shows global maps of O-B difference at three long wave sounding channels, i.e. ch.261, ch.453, and ch.672. We can find cloud distribution in the global map, that means the longwave sounding channels have some information of cloud. Therefore, these three channels are included in 'mix02' and 'sound02' channel set.

<u>Short wave infrared sounding channels and channels</u>

Fig.11 shows global maps of O-B at three short wave sounding channels, i.e. ch.2104, ch.2113, and ch.2333. Sounding channels at shortwave IR region (e.g., ch.2104, 2106, 2108, 2109, 2111, 2113, 2114, 2116, and 2134) are very difficult to use for cloud detection in the daytime because 1) cloud signal at short wave IR channels are smaller than that at short wave IR window channels, 2) sun glint effect is significant even for these sounding channels, and 3) non LTE effect can be seen the upper sounding channels, e.g., ch.2104 in the daytime. Therefore, these shortwave sounding infrared channels are not included in the cloud detection channel set in this study.

Short wave infrared window channels in the daytime

Window channels at short wave IR (e.g., ch.2328 and 2333) are severely influenced by solar irradiance in the daytime as shown in Fig.10, so we cannot use these window channels in sun glint region at all. Also in the dawn region with very low sun altitude, the short wave IR window channels are difficult to detect cloud because the reflectance of cloud and ocean can be almost the same.

However, these short wave IR window channels are very useful for cloud detection other than sun glint region and dawn region because cloud reflectance is much higher than that of ocean. Actually, O-B patterns at ch.2328 and ch.2333 between 10S-50S are well matched with geostationary satellite VIS images.

To confirm the advantages of these short wave infrared window channels, we examine 'mix02' and 'sound02' channel set in this study even in the daytime case.

Microwave channels

Microwave channels, AMSU-A ch.3 and AMSU-A ch.15, are used in the mix02 and sound02 channel sets. AMSU-A ch.2 used in Takeuchi (2002) is removed because the O-B statistics suggests that the channel has large negative bias about 5K for clear case. Fig.12 is the global map of the O-B difference of AIRS-A ch.2. We can find most of O-B diference at clear area has large negative value (violet in the map). Since such large bias disables the correct cloud detection, AMSU-A ch.2 is removed from the cloud detection channel sets.

4.3.4 Background error covariance and thresholds for the Var, PCA, and Optional PCA scheme

The same background error covariance matrix used in the simulation study is applied to the Var scheme to calculate cloud cost. However, the standard deviation of skin temperature is changed from the original value of 2.24K, used in the prior study, to a new value of 0.6K. This change gives remarkable improvements, not shown here, especially for low level cloud and for thin cloud. The same background error covariance matrix is also used for the PCA scheme and Optional PCA schemes.

A cloud cost threshold of 0.94 proposed in the prior study is used both for mix02 channel set and sound02 channel set in the Var cloud detection scheme. Cloud costs are calculated by IASI_1DVAR code (Collard, 2002). This code can treat reduced channel observation data and multi-instrumental observations. This function enables us to perform this study saving computer memory and calculation time.

A threshold of 2.0 is used for the PCA scheme and applied to all used PCA components. Based on results from the simulation study described in section 2.3, the threshold is applied to the partial cloud cost of the first nine PCA components of mix02 and sound02 channel set.

The threshold for the Optional PCA scheme is 0.0. The partial cloud costs (with clear) of the first six PCA components with large eigen values are averaged to obtain a normalized partial clear cost. The cloud cost (with cloud) is calculated from a given O-B statistics for cloudy case. The O-B statistics are calculated by using the Focus day data themselves, because the previous day data are not available in this case.

5. Results and Discussions

5.1 Original feature of O-B difference

5.1.1 O-B map for used channels

Original O-B images for the used channels in the Var, PCA, and Optional PCA schemes are shown in Fig.10. Red and Violet means the absolute value of O-B difference more than 5.0K. We can see cloud distributions in these images. Cloud shows violet appearance in the infrared channels and red appearance in the microwave channels.

The O-B value of cloudy area is lower in the nighttime than in the daytime in the infrared channels. On the contrary, the O-B value is higher in the nighttime than in the daytime in the microwave channels. This diurnal change can not be seen in clear area, therefore, it can be explained by increasing cloud amount and/or thickness in the nighttime.

Secondly predominant features can be seen in the shortwave infrared channels. Very intensive solar signal locates around the center of each scan over the North Hemisphere. The Aqua satellite crosses the equator daily at 1:30 p.m. as it heads north. It means the reflected sun light is seen lefthand of the ascending orbit and the sun glint position is changing in season. For the focus day, i.e. 20 July 2002, the sun glint has the intensity peak around 30degN.

Third feature is a scan angle dependency in the AMSU-A1 channels. The O-B value shows the large negative value in the right hand edge of the direction of satellite. The feature is evident for AMSU-A ch.3. The magnitude of the scan angle dependency is different between ch.3 and ch.15, so cloud detection will be suffered by this instrument oriented anomaly.

5.1.2 PCA results for cloud characterization

To show cloud signal is a predominant component of the original O-B image, in this section, we show PCA images of O-B difference and statistics of O-B data. The results are shown by daytime and nighttime and mix02 channel set and sound02 channel sets, respectively. The PCA values are calculated from the covariance matrix constructed from cloudy O-B values in Focus day data at 00, 06, 12, 18UTC. In these images, the violet and blue color denotes negative PCA values and red color denotes positive PCA values. The enlarged PCA images are also used in the case studies in section 6.

5.1.2.1 Daytime

<u>mix02 channel set</u> (Table 11 and Fig.13-1)

The characteristics of each PCA image with mix02 channel set in the daytime are as follows,

1) PCA01, characterized by positive O-B in the short wave infrared channels, corresponds to reflected solar irradiance intensity. In the North Hemisphere strong sun glint is remarkable between 20deg N to 40deg N. The strong reflection is also found in the high latitude more than 40deg N. In the South Hemisphere, the pattern of PCA01 is well matched to cloud distribution verified by the AIRS VIS/NIR images. Small contribution from surface temperature might be included because the local minimum value each latitude is larger around Tropics and smaller high latitude in the South Hemisphere. Clear ocean in the North Hemisphere shows weak solar reflection at sea surface.
2) PCA02, characterized by positive O-B in the microwave channels and small negative O-B in the infrared channels, mainly corresponds to cloud liquid water. Large value in the higher latitude than 60deg N is due to the poor background surface emissivity. Low PCA02 corresponds to clear area.
3) PCA03 characterized by negative O-B in the microwave channels, negative in the longwave infrared channels, and small positive in the shotwave infrared channels, correponds to high-level ice cloud without underlying low-level water cloud, because ice cloud is insensitive to the microwave but the temperature is lower than that of the surface and it reflects solar radiance in the shortwave infrared region.
4) PCA04, characterized by the difference between AMSU-A ch.15 and AMSU-A ch.3, corresponds to scan angle dependence of these brightness temperatures. Large asymmetric feature and small symmetric feature associated with scan angle can be found. The amplitude of this component is about 1.6K. It should be noted that some meteorological signals, might be water vapor amount, can also be found in the PCA04.
5) PCA05, characterized by positive O-B in the longwave sounding infrared channels and negative O-B in the surface sensed infrared and microwave channels, may correspond to vertical temperature and water vapor profile. The pattern of the PCA05 image does not represent the cloud pattern directly.
6) PCA06-12 have very small amplitude less than 0.7K, therefore, the physical meaning of these components is difficult to explain and the signal is difficult to extract. For example, the PCA12 image is shown in Fig.13-1. Though some signal can be found between 10degN to 20degN in the Atlantic Ocean and the Southern part of the Mediterranean sea, which might be associated with Saharan dusts, the other parts are actually noisy.

<u>sound02 channel set</u> (Table 13 and Fig.13-2)

The PCA images with sound02 channel set is similar to those with mix02 channel set except for the first PCA component with mix02 channel set related to reflected solar irradiance. PCA01-04 components with sound02 channel set have quite similar pattern with PCA02-05 components with mix02 channel set, though PCA03 and PCA04 components with sound02 channel set have inverse signs of those with mix02 channel set. It should be noted that PCA02 component with sound02 channel set is not affected by solar irradiance, therefore, we can cleary see the distribution of high-level ice cloud without sun glint pattern.

5.1.2.2 Nighttime

mix02 channel set (Table 12 and Fig.14-1)

The characteristics of each PCA image with mix02 channel set at night are as follows,

1) PCA01, characterized by small positive O-B in the microwave channels and negative O-B in the infrared channels, mainly corresponds to cloud. Active cloud region associated with a low pressure system west of Europe shows large PCA01 value. Low PCA01 corresponds to clear area.
2) PCA02 characterized by positive MW, small negative IR O-B correponds to low level water cloud as shown around the North East Pacific Ocean. High level ice cloud is also seen in this image because such ice cloud tend to have negative brightness temperature in the infrared region but the cloud cannot be seen in the microwave channels.
3) PCA03 characterized by positive shortwave IR and negative longwave IR corresponds to the phase of cloud particles. Ice cloud has positive PCA03 value and water cloud has negative PCA03 value. Clear area has an intermediate value.
4) PCA04 characterized by the O-B difference between AMSU-A ch.3 and AMSU-A ch.15 corresponds to scan angle dependency of these brightness temperatures similar to the PCA04 for mix02 channel set in the daytime. Large asymmetric feature and small symmetric feature associated with scan angle can be found. The amplitude of this component is about 1.6K.
5) PCA05-12 have very small amplitude less than 0.9K, therefore, the physical meaning of these components is difficult to explain and the signal is difficult to extract. PCA05 seems to relate to the variability of low-level temperature profile and PCA06 relates to high-level temperature profile variability. The PCA12 image shows really noisy and the signal of Saharan dusts can be seen off the coast of Western Africa, but the amplitude is weaker than that in the daytime.

sound02 channel set (Table 14 and Fig.14-2)

In general, the PCA images with sound02 channel set is similar to those with mix02 channel set. The amplitude of PCA01 component of 9.5K for sound02 channel set, corresponds to cloud property, is smaller than that of 10.5K for mix02 channel set. PCA04 related to scan angle dependency of these brightness temperatures has the inverse sign of that for PCA04 with mix02 channel set.

PCA03, characterized by positive O-B in the midium wave (8.9micron) infrared and negative O-B in the longwave infrared, corresponds to the phase of cloud particles. This component is similar to the PCA03 with mix02 channel set, in short, the 8.9 micron window band is used as a substitute for the 3.8 micron band to represent the cloud phase property.

5.2 Cloud detection map (Global feature)

The first result from cloud detection is cloud detection map with the Optional PCA, PCA, Var, Mitch, and ECMWF schemes as shown in Fig. 15-1 (daytime) and 15-2 (nighttime). For the Optional PCA, PCA, and Var scheme, a couple of results are shown for mix02 channel set and for sound02 channel set, respectively. Blue color means clear, green for cloudy. Red means FOVs for larger cloud cost than 20 for the Optional PCA,, PCA, and Var scheme.

5.2.1 Daytime case (Fig.15-1)

In the daytime, the Optional PCA, PCA, and Var scheme with mix02 channel set are strongly affected by solar reflection, then, these schemes detect almost no clear area in the North Hemisphere. In the South Hemisphere, these schemes detect some clear area. The position of the clear area is consistent among these scheme, but the Optional PCA scheme is more strict than the other two schemes. The Var scheme gives more clear area than the other schemes. Mitch's scheme, which uses a shortwave infrared channel, is also affected by solar reflection. This scheme gives much more clear area but the distribution of the declared clear area is not consistent with the other schemes.

The Optional PCA, PCA, and Var scheme with sound02 channel set and ECMWF scheme are not affected by solar reflection because these schemes do not use shortwave infrared channels. Though the result of Var scheme with sound02 and ECMWF scheme is similar, the Var scheme finds more cloud in the high latitudes and the ECMWF scheme finds more cloud in the low latitudes. This latitude dependency is consistent to the result of the simulated study. Because the ECMWF scheme uses single band channels, it cannot detect ice cloud efficiently. On the other hand, the Var scheme uses multi-band channels, in the daytime 8.9 micron bands plays an alternative role to the 3.8micron bands, therefore, it can detect ice cloud at high-latitudes efficiently.

Unfortunately, the Var scheme and the ECMWF scheme detect not as much cloud as expected in the visible images of geostationary satellite and the AIRS VIS/NIR photometer. The PCA scheme with sound02 channel improves the result but it is not sufficient. Optional PCA scheme with sound02 channel seems best because it detects more cloud even in low- and mid- latitudes.

5.2.2 Nighttime case (Fig.15-2)

In the nighttime, the Optional PCA, PCA, and Var scheme with mix02 channel set work well due to the absence of solar reflection. The Optional PCA scheme is the most strict and Var scheme gives more clear areas as in the daytime case. In this case the Var scheme with mix02 channel set, that with sound02 channel set, Mitch's scheme and ECMWF scheme all give similar result. The difference between mix02 channel set and sound02 channel set is not clear for the PCA scheme and Var scheme. On the other hand, the difference is significant for the Optional PCA scheme. It is because the cloud probability is better represented by including the shortwave infrared channels.

5.2.3 Partial cloud costs in the Var and PCA scheme

To examine what PCA component has the maximum contribution in cloud detection with PCA and Var schemes; the partial cloud costs are plotted against PCA indices for all (clear and cloudy) FOVs. Fig.16-1 is the result in the daytime. In the cloud detection with mix02 channel set, PCA07 to PCA10 have extremely large contribution. For these PCA components, background variance is very small, however, the O-B difference of the shortwave infrared channels is relatively much larger than that of the background due to solar reflectance. Then, the PCA components with small eigenvalues play important roles. On the other hand, in the night case, PCA02 and PCA04 are effective in the cloud detection. In these components, cloudy variability is large compared to clear variability. For PCA05 to PCA09 have medium sensitivity to cloud but occasionally large partial cost function are calculated. PCA 10 to PCA12 have weak sensitivity to cloud.

For the sound02 channel set, the feature is similar in the daytime and nighttime. As the results of mix02 channel set in the nighttime, the second and fourth components are sensitive to cloud, and over fifth components have medium sensitivities but ocasionally large partial cost function are calculated.

5.2.4 Effect of the truncation of PCA index in the PCA scheme

To examine the truncation effect of PCA index in the PCA scheme, each cloudy FOV is classified by the PCA index having the maximum partial cloud cost. In Fig. 17-1 and 17-2, green area is declared as cloudy by both of the first six PCA components and the 7^{th} to 9^{th} PCA components. The yellow area is declared as cloudy only by the first six components, and the red area is declared as cloudy only by the 7^{th} to 9^{th} components. In the daytime case in Fig.17-1, Red FOVs distribute not systemtically as these higher components play only a secondary role. Also in the nighttime case, the red FOVs are fewer than green and yellow. It can be concluded that the truncation of the higher PCA components does not degrade the performance of cloud detection.

It should be noted that the yellow area is widely distributed in the low- and mid-latitudes, particularly for mix02 channel set in the nighttime. It suggests that the higher components are insensitive to cloudy areas in low- and mid-latitudes. This result agrees with the result of cloud detection with the Var scheme which does not work well in low- and mid-latitude in Takeuchi (2002).

5.2.5 Day/night difference of cloud probability distribution used in the Optional PCA scheme

In the Optional PCA scheme, cloud probability should be given by day and night, separately. Fig.18-1 shows a cloud detection map with mix02 channel set in the daytime, one with nighttime cloud probability, and the other with daytime cloud probability. The clear area is less with daytime probability than with nighttime probability. This result suggests that the cloud probability distribution is different between daytime and nighttime.

The difference of the cloud detection performance is more remarkable in the case of sound02 channel set. In Fig.18-2, the performance of cloud detection in the daytime with the Optional PCA scheme is much improved by applying the daytime cloud probability distribution.

5.3 O-B statistics for clear FOV (Global statistics)

In this section, O-B statistics for clear FOVs are investigated on global basis.

Table 7-10 show PCA components of clear FOV, where clear FOVs are declared with the PCA scheme for mix02 and sound02 channel set and in the daytime and nighttime.

5.3.1 Mean and standard deviation of O-B difference of clear FOV (Fig.19 and Table 5)

In Fig.19 shows mean and statndard deviation of O-B difference of clear FOV for all distributed AIRS and AMSU-A channels in the daytime. Since the three schemes with sound02 channel set and the ECMWF scheme do not use shortwave infrared channels, large positive O-B difference due to solar radiation is observed. Additionally, the three schemes with sound02 channel set shows smaller standard deviation in 8.9 micron bands than the ECMWF scheme. It suggests that the combined use of multi-bands channels is efficient.

On the other hand, the three schemes with mix02 channel set and the Mitch's scheme use shortwave infrared channels, so O-B difference in the shortwave infrared band is very small. However the O-B difference has a small positive bias in the band, therefore, it suggests that the solar radiation effect is not removed completely by the schemes and the thresholds used.

In the nighttime case in Fig.20, the O-B difference of window channels are very small for all schemes. The ECMWF scheme, not using shortwave infrared channels, shows a slightly large standard deviation in the shortwave infrared channels. The Optional PCA scheme seems to give the smallest standard deviation of clear FOVs among all the schemes.

5.3.2 O-B histogram for selected channels

Fig.21-1 shows the O-B histograms for selected channels in the daytime. The Optional PCA, PCA, and Var scheme is of the sound02 channel set. Red lines for clear FOVs and green lines for cloudy FOVs. A feature of the ECMWF scheme is that the scheme divides clear FOVs and cloudy FOVs by longwave infrared channels, then cloud with small O-B difference at ch.914 are hardly detected by the scheme. On the other hand, the Var, PCA, and Optional PCA schemes can find some cloud FOVs with small O-B difference at all channels. This advantage is remarkable in the PCA and Optional PCA scheme.

Fig.21-2 is the result for nighttime. The feature is similar to the daytime case. The ECMWF scheme has clear FOVs with large O-B value at AIRS ch.2333 and AMSU-A ch.3. Asymmetric distribution can be found for the O-B histogram at AMSU-A ch.15, it might be due to the assumption of cloud probability distribution used here does not represent the true distribution completely.

5.3.3 Dependency of O-B statistics to thresholds

In this section, we investigats the dependency of O-B statistics to thresholds used in each scheme. Mean (bias) and standard deviation of O-B difference for clear FOVs is shown for AIRS ch.914, ch.2333, and AMSU-A ch.3 in Fig.22. Since the thresholds used in each scheme have different meanings, the performance of the cloud detection schemes cannot be compared to each other by the thresholds, then, the numbers of the detected clear FOVs are used in the horizontal axis in this comparison. Number of clear profiles is an indicator of strictness of cloud detection. Fewer clear profiles means that the scheme and the given threshold are more strict. Of course, the cloud distribution by each scheme should be validated by other data such as high resolution imager data as described in section 6.

Mean values of O-B is an indicator of cloud detection performance because undetected cloud makes the mean value negative in the infrared region and positive in the microwave window region in general. The Var and PCA scheme give small absolute mean values.

By normalizing with the clear data number, the PCA scheme has almost the same mean and standard deviation as Var scheme. The Optional PCA scheme has different mean values from the other schemes and the mean value is near to the mean value with the PCA and Var scheme with half the number of clear FOVs derived from the Optional PCA scheme. It means that more data can be used in the assimilation with a given bias correction. The standard deviation of the Optional PCA scheme is slight better than the other schemes for thresholds giving clear FOVs more than 3000, i.e. 8% of all FOVs,

O-B statistics of Mitch's scheme and ECMWF scheme show that the thresholds in these schemes are too loose to obtain clear profiles for bias calculation and/or retrievals. In addition, these schemes have less performance than the Var, PCA, and Optional PCA scheme because the Mitch's and ECMWF schemes show much more cloud contamination (i.e. large negative bias at ch.2333 and large positive bias at AMSU-A ch.3. For ch.914 ECMWF scheme has almost the same performance as Var, PCA, and Optional PCA schemes but Mitch scheme shows large negative bias due to not using background temperature and water vapor information.

5.4 Calculation time - a key issue to realize the scheme on an operational processing -

Execution time on a HP workstation was estimated for cloud detection with each scheme by inserting date command just before and after the IASI_1Dvar command. Table 6 is the total exection time to process 10 profiles. The execution time for 4001st to 4010th profiles at 00UTC 20 Jul 2002 are shown. The results show the PCA and optional PCA schemes require almost no additional the execution time. The reason that exection time of ECMWF scheme is more than those of other schemes is that the ECMWF scheme needs forward calculations for all channels, i.e. 339ch.

6. Results and Discussions (Case study)

6.1 Common discriptions

6.1.1 Purpose of case studies

The purpose of case studies is the following.

- To make clear advantages and shortcomings of cloud detection schemes in different synoptic situations.
- To make clear the detailed performance of cloud detection schemes by comparing with associated high resolution visible and near infrared imager data, namely AIRS VIS/NIR images.

The cases are selected aiming at,

- The case includes clear and cloudy areas with rather distinct features in geostationary satellite browse images on the Focus day.
- Different synoptic situations, such as with lower cloud, high thin cloud in the area, or solar insolation affected case.

6.1.2 Figure description used in this case study

6.1.2.1 Daytime

To compare the results of the cloud detections, sectionalized cloud maps with Optional PCA, PCA, Var, Mitch, and ECMWF schemes are shown. Used channels used in the Optional PCA, PCA, and Var schemes are mix02 channel set sound02 channel set. Blue color shows clear area and Green color shows cloudy area. The red color shows the area the cost of which is more than 20.0.

The first to fifth PCA components calculated from whole O-B data at daytime with sound02 channel set, whose global result is discussed in section 5.1.2, are shown. Here, the mix02 channel set is not used because the major PCA components with mix02 channel set consists of contribution from solar irradiance and do not represent cloud contribution. The first PCA component has characteristics of positive MW and negative IR related to general cloud property. The second component is characterized by negative MW and negative IR values related to high-level ice cloud without underlying low-level water cloud. The third component is related to scan angle dependency of the MW channels. The fourth component is related to vertical temperature and water vapor profiles.

Since the simultaneous AIRS VIS/NIR channel images coincident with AIRS infrared channels are only available at daytime, the validation is more strict than that in the nighttime cases. Enhanced images are also shown to take a look at the thin cloud distribution.

6.1.2.2 Nighttime

As for daytime case, the same kind of sectionalized cloud map are shown. The results with mix02 channel set are shown for the Optional PCA, PCA, and Var schemes to confirm effectiveness of the near infrared channels.

To compare the results of the cloud detections, sectionalized cloud maps with Optional PCA, PCA, Var, Mitch, and ECMWF schemes are shown. Used channels used in the Optional PCA, PCA, and Var schemes are mix02 channel set. Blue color shows clear area and Green color shows cloudy area. The red color shows the area the cost of which is more than 20.0.

To assist in the interpretation of the comparison, the first to fifth PCA components calculated from whole O-B data of mix02 channel set at nighttime are also shown. The redish colors show high PCA values and the violet color shows low PCA values. The first component have characteristics of small positive MW and negative IR. In the images, clear area has bluish color, thick ice cloud and water cloud has redish color. The second component consists of positive MW and small negative IR values. In this image, low level water cloud and high level ice cloud are seen. The third component have positive shortwave IR and negative longwave IR related to the phase of cloud particles. In this image, violet area means water cloud and redish area means ice cloud. The fourth component shows microwave scan angle dependency, then, it is an instrument origin component. The scan edge shows more redish and scan center is more violet. The fifth component consists of positive O-B in shortwave IR and longwave IR sounding channels and negative IR window channels. This components reflects some of vertical profile differences.

Since AIRS VIS/NIR false composite images described in section 4.3.3 are not available at night, the daytime images on the same day are shown giving a time-lag of about a half day and the coverage difference. The enhanced images are also shown to identify very thin cloud distribution. It should be noted that we cannot verify the individual cloud distribution but synoptical cloud type and distribution.

6.2 Daytime cases

6.2.1 North coast of Australia (Fig.23)

In this area (30S-0S,110E-150E), low-level water cloud band north west of Australia can be seen in PCA01 image and AIRS VIS/NIR image. Gulf of Carpentaria seems to be clear but the enhanced AIRS VIS/NIR image shows that very thin cloud or haze is extended off coast.

Optional PCA and PCA scheme with mix02 channel set detect not only the lower water cloud band but also detail featured cloud rows near the coast. Var scheme with mix02 channel set also detects the cloud band well and extracts the pure clear area in the Gulf of Carpentaria very well. Optional PCA scheme with sound 02 channel set also succeeds in finding both of low-level cloud and very thin cloud. PCA scheme with sound02 channel set detects only rather thick water cloud in the cloud band north-west of Australia. Var scheme with sound02 channel set does not detect the cloud band at all. Mitch's scheme gives a doubtful round cloud area north west coast of Australia, where the above three schemes declare clear. ECMWF scheme also gives a different cloud distribution from that of other schemes and the distribution is not consistent to AIRS VIS/NIR image. The best choice for this case is the Optional PCA scheme with mix02 channel set. This case apparently shows the advantage of Optional PCA scheme over the other schemes and the importance of use of near infrared channels in the daytime and/or cloud probability distribution.

Fig.23-5 shows the impotance of cloud probability dependent of daytime and nighttime. Cloud distribution with daytime cloud probability gives reasonable, however, that with nighttime cloud probability gives much clear area which is cloudy in AIRS VIS/NIR image.

Fig.23-6 is a result after adjusting thresholds so that the number of clear FOVs declared by each scheme, i.e. Optional PCA, PCA, and Var scheme, is about 1050, i.e. 12% of all ocean FOVs. The result is similar each other after the adjustment, however, the PCA and Var scheme gives noisier cloud pattern. On the otherhand, Optional PCA pattern gives systematic cloud pattern.

Fig.23-7 shows the dependency of cloud detection maps with Optional PCA scheme to threshold. The threshold of 0.0, corresponds to total clear fraction of 12%, gives very strict cloud detection and the threshold of 0.10, corresponds to total clear fraction of 20%, gives a relaxed cloud detection. It is noted that the cloud pattern is consistent when the threshold is changed from 0.0 to 0.10. We can modify the threshold to get the largest performance in retrievals and/or assimilation of AIRS/AMSU-A.

6.2.2 South Africa and Mozambique Channel (Fig.24)

In this area (40S-10S,10E-50E), lower water clouds distributed in Mozambique Channel and south western sea of South Africa can be seen in the AIRS VIS/NIR images. Pure clear area can be seen to the west coast of Madagascar and the south-west coast of South Africa. As in the previous case, most of clear area is covered by very thin clouds.

In Optional PCA scheme almost all FOVs are declared as cloudy and the residual clear area is not coincided with that in AIRS VIS/NIR image. PCA scheme with mix02 channel set gives the reasonable clear area in Mozombique Channel. Optional PCA scheme with sound02 channel set also gives a good result. PCA scheme with sound02 channel set and Var schemes with both channel sets give less cloud than the above schemes. However, Var scheme with mix02 channel set looks better than the others for the cloud distribution over south-western coast of South Africa. Mitch's scheme gives bad results uncorrelated with AIRS VIS/NIR image. ECMWF scheme gives less cloudiness over this region.

These results also show the use of short wave infrared channels is good for accurate cloud detection. The best choice for this case is the PCA scheme with mix02 channel set. Optional PCA scheme with sound02 channel set is of the second choice.

6.2.3 Sea around Japan (Fig.25)

In this area (20N-50N,110E-150E), an intensive sun glint area can be seen around left-hand of center of each scan. The three check points in the Japan sea are; 1) dense ice cloud along northern part, 2) low level cloud in the southeast part, and 3) very thin cirrus cloud east coast of Korea peninsula, which corresponds to eastern edge of widely spreading clouds over Yellow sea. South of Japan are widely covered cumulus type clouds associated with medium or high level very thin cloud, therefore, almost no clear area can be found in this area.

The Optional PCA, PCA, and Var schemes with mix02 and Mitch scheme are affected by the solar insolation while other schemes, i.e. three schemes with sound02 channel set and ECMWF scheme, are not affected by the solar insolation.. The Var scheme with sound02 channel set and ECMWF scheme do not detect most of the lower clouds in this area. The PCA scheme with sound02 channel set gives rather better result. The Optional PCA scheme with sound02 channel set detect very thin cirrus east coast of Korea peninsula and low level cloud in the southeast of Sea of Japan. Cloud detection in southern sea of Japan is succeeded only by this scheme.

These results shows the use of cloud probability distribution is essential to find very thin cloud. When we need a possible clear FOV even in sun glint area, the best choice is the Optional PCA scheme with sound02 channel set. In the other cases, the PCA scheme with mix02 channel set is the best one.

6.2.4 Western coast of Europe and Mediterranean Sea (Fig.26)

Also in this area (20N-50N,30W-10E), the intensive sun glint area can be seen. A low pressure system west of Spain has many curved cloud bands around the center. West part of Mediterranean sea looks clear at a glance. But the coastal area of North Africa and Spain are partly coverd by Saharan dust and thin ice cloud as identified by AIRS VIS/NIR image. The Saharan dust also widely covers the Atlantic Ocean less than 35deg N.

The Optional PCA, PCA and Var schemes with mix02 channel and Mitch scheme are affected by the solar insolation as well as the above case. The Var scheme with sound02 channel set and ECMWF scheme give similar results. These schemes look over the Saharan dust over the Atlantic Ocean. The Optional PCA and PCA scheme with sound02 channel set find more cloud than other schemes, particularly in the Mediterranean sea. Saharan dust over Atlantic Ocean is detected by Optional PCA scheme with sound02 channel set. In this case, the Optional PCA scheme with sound02 channel set gives the best results.

Fig.26-5 is a result after adjusting thresholds as in Fig.23-6. Though the difference between the schemes are not significant but it should be noted that clear area off the west coast of France is detected only by Optional PCA scheme.

Fig.26-6 shows the dependency of cloud detection maps with Optional PCA scheme to threshold as in Fig.23-7. The threshold of 0.10 gives more cloud over the Atrantic Ocean and the Mediterranean sea than the threshold of 0.0.

6.3 Nighttime case

6.3.1 NorthEast Pacific Ocean (Fig.27)

In this area (30N-60N, 150W-110W), low level cloud is widely distributed over the north-east part of the Pacific ocean. The northern area higher than 47degN and some other parts are covered by high-level ice cloud well seen in GOES IR image (not shown here). Some distinct clear part can be found in AIRS VIS/NIR images, the northern part of which is covered the high-level cloud. An enhanced AIRS VIS/NIR image suggests very thin cloud distributes along west coast of U.S.A..

Mitch scheme gives much different results from other schemes, especially it misses distinct large clear areas around 45N142W and 39N148W. ECMWF scheme finds some clear area identified by AIRS VIS/NIR images, but it seems larger than the AIRS VIS/NIR image, especially the scheme cannot detect high thin cirrus cloud and very thin cloud off coast of California. Var scheme gives more reasonable clear area. Optional PCA and PCA schemes give more strict clear area with less cloud fraction. Best choice for this case is the Optional PCA scheme with mix02 channel set.

6.3.2 Mediterranean sea and West coast of Europe (Fig.28)

In this area (30N-60N, 20W-20E), the similar area in section 6.2.4 but a half day earlier than that, a low-pressure system with systematic cloud is located west of Spain. West part of Mediterranean Sea looks like clear. But coast area of North Africa and Spain are partly covered Saharan dust and thin ice cloud discussed in section 6.2.4. The English channel to the North Sea is also covered cloud associated with a small low-pressure system over the Britain Isles.

Var scheme and ECMWF scheme give much clearer area over the Mediterranean Sea and over the English channel. Mitch scheme declares most of western part of Mediterranean is cloudy. Optional PCA and PCA scheme find cloudy area over the Mediterranean Sea and the English channel.

It is difficult to determine the best scheme for this case because the AIRS VIS/NIR images are of different time and coverage. We can only say that the Var and ECMWF schemes seem to give more clear area than truth and Mitch scheme seem to give more cloudy area than the truth over the Mediterranean sea. The PCA and Optional PCA schemes seem to give a reasonable clear area.

6.3.3 North-west Pacific Ocean (Fig.29)

In this area (20N-50N,130E-170E), thick clouds distributed zonally between 37N and 43N and other thick cloud band can be seen between 45N and 49N. No clear area is between these cloud bands. Cumulus type low-level clouds are scattered over most of south east of Japan. Very thin ice cloud is also widely spread over the scattered low level cloud, which is detected as a red area in the enhanced AIRS VIS/NIR image. The enhanced AIRS VIS/NIR image suggests scattered pure clear areas can be found between 30-36N.

All schemes detect the scattered low level cloud over south east of Japan. Var scheme detects less low level clouds than other schemes do. ECMWF scheme does not detect very low level cloud around 46N-50N 163-165E. Optional PCA and PCA schemes detect very thin ice cloud over south east sea of Japan and the Optional PCA scheme gives a reasonable clear area.

6.3.4 Sea around Indonesia (Fig.30)

In this area (10S-20N,80E-120E), no systematic cloud can be seen over all region. Scattered cumulus type cloud and anvil type high- or middle-level cloud, which are associated with active convective cloud system, can be seen in the GMS IR images (not shown here) and the AIRS VIS/NIR images. However light gray region in the GMS IR image and red region in the enhanced AIRS VIS/NIR image suggest that very thin ice cloud covered almost all this area.

All schemes detect the scattered low level cloud over this area. Var and ECMWF schemes do not detect such clouds so much. These two schemes and Mitch's scheme do not detect the high- and middle-level thin cloud. Optional PCA and PCA schemes detect not only scattered low cloud but also very thin cloud. Best choice for this case is the Optional PCA with mix02 channel set.

7. Conclusions

7.1 Summary of results

This report describes new cloud detection schemes refered to as PCA scheme and Optional PCA scheme based on a principal component analysis of AIRS/AMSU-A channel radiances onboard theAQUA satellite. A comparison of these schemes is made with the VAR cloud detection scheme based on a Bayesian theory, Mitch's scheme not using background information, and the ECMWF scheme not using multi-band information.

The AIRS Focus day data are used to compare the performance of these cloud detection schemes and AIRS VIS/NIR high resolution images are used for the validation of these schemes.

The PCA images of the original O-B data of mix02 channel set and sound02 channel set show, 1) the major components relate cloud, its phase, and scan angle dependency of AMSU-A channels, and 2) the predominant component is of solar reflection with mix02 channel set in the daytime.

From this study, the following results can be infered:

1) In the daytime, the Optional PCA, PCA, and Var scheme with mix02 channel set and the Mitch's scheme are strongly affected by solar reflection, and so, these schemes detect almost no clear area in the North Hemisphere in July. In the Southern Hemisphere, these schemes detect some clear areas. Though the Var scheme with sound02 channel set and the ECMWF scheme give similar results, the Var scheme finds more cloud in the high latitude and the ECMWF scheme finds more cloud in the low latitude due to the difference of channels used. Only the Optional PCA scheme with sound02 channel detects much cloud even in low- and mid-latitude.
2) In the nighttime, the Optional PCA scheme is the most strict and Var scheme gives more clear areas as in the daytime case. The difference between mix02 and sound02 channel set is remarkable for the Optional PCA scheme because the cloud probability is well represented by including the shortwave infrared channels.
3) The truncation of the higher PCA components above seven in the PCA scheme does not degrade the performance of the cloud detection.
4) The higher PCA components are insensitive to cloud in low- and mid-latitudes. This result explains why the cloud detection with Var scheme does not work well in these latitudes.
5) In the Optional PCA scheme, the use of the different cloud probability distributions by day/night is essential for accurate cloud detection especially for the case of the sound02 channel set.
6) The results of Var, PCA and Optional PCA schemes are similar when the threshold is adjusted so that the number of clear FOVs declared by each scheme is a constant. But the cloud pattern with Optional PCA pattern is the most systematic and consistent to the cloud pattern in the AIRS VIS/NIR image.
7) The Optional PCA scheme can be tuned by changing the threhold value. We can modify the threshold to get the largest performance in retrievals and/or assimilation of AIRS/AMSU-A. The threshold of 0.10 corresponds to clear fraction of 12% and is recommended as a candidate for the retrievals and assimilation.

The O-B statistics for clear FOVs shows,

8) The ECMWF scheme derives clear FOVs and cloudy FOVs by using the longwave infrared channels, and so, cloud with small O-B difference at the longwave infrared channels are hardly detected by the scheme.
9) By normalizing with the clear data number, the PCA scheme has almost the same mean and standard deviation as the Var scheme. The Optional PCA scheme has different mean values from the other schemes and the mean value is near to that with the PCA and Var scheme with half the number of clear FOVs derived from the Optional PCA scheme.

From the case studies with variable synoptic situations,

10) In the daytime case without sunglint, the shortwave infrared channels are useful to detect cloud due to the contrast of solar reflection of cloud and ocean. The derived cloud pattern is similar to that of AIRS VIS/NIR image. The Optional PCA scheme with mix02 channel set is the best combination.
11) In the daytime case with sunglint, the shortwave infrared channels are no use. The Optional PCA scheme with sound02 channel set with longwave infrared and microwave channels are the best combination.
12) In the nighttime case, the Optional PCA scheme with mix02 channel set is the best choice.
13) Mitch's scheme can be applied both in the daytime and the nighttime, however, the number of the detected cloudy FOVs is less than the other scheme and the scheme occasionally gives some suspect cloud distributions.
14) The ECMWF scheme gives more clear areas mainly due to a loose threshold on the gradient of the filtered O-B value to the 'ranked channel index'. When a tight threshold is given, the scheme gives noisy cloud distributions.

The problem that the lower sensitivity of the multi-channel Var scheme to cloud in mid- and low-latitude than single channel Var scheme is solved by introducing the PCA scheme. The Optional PCA scheme gives more symmetric statistics of observation minus background brightness temperatures by considering the proper cloud probability distributions of cloudy profiles.

7.2 Recommendation for AIRS day2 cloud detection scheme

Optional PCA(mix02) in the nighttime and daytime except for sun glint area and dawn area and Optional PCA (sound02) in the daytime for sun glint area and dawn area

In this study, the efficiency of the simultaneous use of 3.8 micron channels and microwave channels with longwave infrared channels to detect cloud is clearly shown even in the daytime area except for sun glint area and dawn area. Sun glint area is easily identified by sun-earth-satellite geometrical calculation and dawn area is identified by sun-earth geometrorical calculations, therefore, we can use mix02 channel set in the daytime region by adding this information.

A small problem is the continuity the results between mix02 and sound02 channel sets. Of course, 3.8 micron channels cannot be used in 1Dvar retrieval steps in the daytime because the forward model is not applicable for solar radiation. If we want to avoid the discontinuity of cloud detection in the daytime, the use of sound02 channel set in the all daytime data is the alternative option. As shown in section 6, the performance of Optional PCA scheme with sound02 channel set is almost the same as that with mix02 channel set by applying the different cloud probability distributions by daytime and nighttime.

The threshold of 0.10 in the Optional PCA scheme recommended as a candidate for the retrievals and assimilation, but the optimum threshold should be found through assimilation experiments.

8. General remarks and suggestions for future work

Including AIRS VIS/NIR data in cloud detection

As shown in section 6, AIRS VIS/NIR image and its enhanced version used as a validation data are very useful to find cloud in the daytime. AIRS VIS channels are particulary powerful to detect low thin cloud and widely distributed very thin ice cloud. The VIS ch.3 of these four channels of AIRS VIS/NIR photometer is the most useful to detect such clouds in combination with AIRS VIS ch.1 and ch.2. These channels should be included in the near real-time AIRS data.

As requested from the user group, NESDIS has included the visible information co-located to the AIRS footprints in the AIRS BUFR data (AIBT BUFR data) distributed to the U.K. Metoffice since 22 January 2003. The values that will be included are: the Cloud Fraction at each AIRS footprints determined by the four visible channels (currently JPL is putting a 20% error on this number), the top of atmosphere Albedo for each Visible channel averaged over each AIRS footprint, the Standard Deviation of the top of the atmosphere Albedo for each Visible channel at each AIRS footprint, and the Visible channel/frequency information (Wolf, 2003).

This additional data will give much improvement to cloud detection in the daytime, though some consideration is necessary in the sun glint area.

Correcting bias of forward model calculation (model and/or input data) and its scan angle dependency.

Cloud detection schemes based on Bayesian theory are very powerful because they can consider the total measurement error covariance and background error covariance in these scheme. These schemes detect the signals originating from cloud, solar irradiance, instrumental noise not included in the forward model from the signal of clear areas. On the other hand, the performance of this scheme relies on the accuracy of the forward model, therefore, the bias estimation and correction of the forward model is a key issue to improve the performance of cloud detection.

In this study, some biases are found in ozone band, water vapor band, 4.2 micron upper sounding band, and 4.2 micron window band, and AMSU-A channels. The first priority to be sorted out is on the scan angle dependency of bias of AMSU-A channels. This bias is a large PCA component of original image as shown in section 5.1.2. In some cases in section 6, problems can be found. Since this scan angle dependency is due to no anntena sidelobes correction having not been applied yet as in section 4.1, the anntena correction should be implemented in the original BUFR data by NESDIS.

The correction of the large biases in water vapor band and 4.2 micron upper sounding band is a secondary priority. This correction should be implemented in doing the cloud detection per channel. The ECMWF scheme, a scheme of the cloud detection per channel, does not work well when such biases are embedded in the original brightness temperature data.

Remarks on the Background error covariance matrix

The background error covariance used in the Optional PCA, PCA, and Var scheme is related to the expected error in a 6hr NWP forecast. A background error covariance matrix calculated for Mid-Latitude Winter is applied to cloud cost calculation in this study. The constant matrix is given both for all latitudes. Considering the dependency of background error to latitude, or synoptic state should give better results.

Remarks on the probability distribution of O-B in the cloud contaminated region

In the Optional PCA scheme, the probability distribution of O-B might be different for each cloud type and cloud height and also background temperature/water vapor profiles. The global mean feature is used in this study but considering the dependency will give better results for polar or tropical regions.

Receiving of Near-real time AIRS data and the quality monitoring

Near-real time AIRS data (AIBT BUFR data) is being received from NESDIS CEMCYS computer via the dedicated link to UK MetOffice and stored in the MetDB since 11 October 2002. Since 19 December 2002, the AIRS data monitoring site was opened. The site is very useful to check some statistical features and their geographical locations. Some investigations with near real-time AIRS data will be used to confirm the result of this study and to fix the Day 2 algorithm.

The cloudy case is predominant in actual profiles, then, clear area obtained by the Optional PCA scheme, which gives more symmetric O-B statistics for the same number of clear FOVs, is effective to assimilation of AIRS data. This will be confirmed in an Observation System Experiment.

Cloud detection per channel

Cloud detection per channel is a key issue to utilize all channels unaffected by cloud in the NWP application. To realize cloud detection per channel, the existing Bayesian cloud detection scheme should be extending the variable channel set, in other words, the ECMWF scheme should be extended so as to treat the multi-bands simultaneously.

This cloud detection per channel is also required to utilize the AIRS data over land and sea ice, where the window channels are affected by surface temperature variability.

The ECMWF scheme requires the forward calculation of all AIRS channels. If the channels are thinned, the cloud detection performance is much reduced because the low-pass filter and thresholding in the scheme does not work well. The reduction in computer time is the key issue to be considered when forward model calculations consume much computer time.

9. Acknowledgements

This study was carried out under the grant of NWP-SAF program. The author presents his sincere acknowledgements to, Roger Saunders for valuable suggestions and arrangement and encouragement of this mission, Andrew Collard for valuable suggestions, providing background and pre-processed observation data, Steve English for valuable suggestions, James Cameron for monitoring the near real time AIRS data, Jonathan Smith for giving the information on AIRS VIS/NIR data and the collocated MODIS data, Pete Francis and Peter Rayer for their technical assistance, Tony McNally, Phil Watts, and Jean-Noel Thepaut for giving important information about the ECMWF scheme, Frederic Chevallier for constructing ECMWF profile database. The author also presents his deep acknowledgements to John Eyre, John Armstrong, and Bryan Conway for arranging this Visiting Science mission.

References

Chevallier,F., 2001: Sampled databases of 60-level atmospheric profiles from the ECMWF analyses, EUMETSAT/ ECMWF SAF programme, Research Report, 4, 27pp.

Collard,A.D., 2002: NWP SAF 1DVar User Manual version 2.3, NWPSAF.

Collard,A.D. and S.B. Healy, 2002: The combined impact of future space-based atmospheric sounding instruments on numerical weather prediction, *Q.J.R.Meterol.Soc.*, **128**, 1-16.

English,S.J., J.R.Eyre, and J.A.Smith, 1999: A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction, Q.J.R.Meteorol.Soc., 125, 2359-2378.

English,S.J. and T.J.Hewison, 1998: A fast genetic millimetre wave emissivity model. *Microwave Remote Sensing of the Atmosphere and Environment Proc. SPIE* **3503**, 22-30.

Goldberg,M., and L.Zhou, 2002: AIRS clear detection flag (Presentation material at a meeting).

Kidwell,K.B eds., 1998: NOAA Polar Orbiter Data User’s Guide. November 1998 Revision., NOAA/NESDIS.

Kinne,S. and K.N.Liou, 1989: The effects of nonsphericity and size distribution of ice crystals on the radiative properties of cirrus clouds. *Atmos. Res.*, **24**, 273-284.

Kneizys,F., J.Chetwynd, R.Fenn, E.Shettle, L.Abreu, R.McClatchey, W.Gallery, J.Selby, 1980: Atmospheric Transmittance/ Radiance Computer Code LOWTRAN 5. Scientific Report, AFGL-TR-80-0067, Air Force Geophysics Laboratory, 233 pp.

Saunders,R.W., P.Brunel, F.Chevallier, G.Deblonde, S.J.English, M.Matricardi, and P.J.Rayer, 2002: RTTOV-7 – Science and Validation Report. *NWP Forecasting Research Tech. Rep.*, 387, 51pp.

Sherlock,V., 1999: ISEM-6: Infrared Surface Emissivity Model for RTTOV-6. NWP-SAF report.

Stephens, 1984: The parameterization of radiation for numerical weather prediction and climate models. *Mon. Wea. Rev.*, **112**, 826-867.

Atmospheric Dynamics DST, 2002: AIRS/AMSU/HSB July 20, 2002 Sample Day L1B Data Documentation, NASA/GES DAAC/Atmospheric Dynamics Data Support Team, 17pp.

Harris,A., 2002: AIRS Version 2.2.3 Processing Files Description, JPL D-20001, Jet Propulsion Laboratory, California Institute of Technology, April, 2002.

Takeuchi,Y., 2002: Cloud detection by using cloud cost for the Advanced Infrared Radiometer Sounder (Part I) - A simulation study -, 22pp. (internal document of UKMetOffice).

Watts,P., T.McNally, M.Matricardi, F.Chevallier, J.N.Thepaut, 2002: Cloud detection for AIRS data and prospects for CO2 estimation, Meteorological applications of data from high resolution spectromenters, 18 April 2002, (Presentation material at the meeting).

Figure and Table captions

Table 1 AIRS channels and the central wave numbers for 324ch selected channels
Table 2 Cloud detection channel set
Table 3.1 Principal Component Analysis for pure clear case with mix02 channel set and statistics and PCA for cloudy case in the PCA space.
Table 3.2 As Table 3.1, but for sound02 channel set.
Table 4 Summary of cloud detection schemes
Table 5 Statistics of clear O-B difference for ocean night case
Table 6 Execution time (unit: sec) for cloud detection of 10 profiles
Table 7 PCA values in the clear case for ocean daytime with mix02 channel set
Table 8 PCA values in the clear case for ocean nighttime with mix02 channel set
Table 9 PCA values in the clear case for ocean daytime with sound02 channel set
Table 10 PCA values in the clear case for ocean nighttime with sound02 channel set
Table 11 PCA values in the cloudy case for ocean daytime with mix02 channel set
Table 12 PCA values in the cloudy case for ocean nighttime with mix02 channel set
Table 13 PCA values in the cloudy case for ocean daytime with sound02 channel set
Table 14 PCA values in the cloudy case for ocean nighttime with sound02 channel set

Fig.1 Illustrating the cloud detection study with real AIRS/AMSU data.
Fig.2 PCA components of O-B difference simulated for ECMWF profile dataset for mix02 channel set. Horizontal axis is for the first component and vertical axis for the another component. Red plots are for clear case and green for cloudy and total liquid water excess total ice water case, and blue are for cloudy and total ice water excess total water cloud case. Blue contours are mean ice water content (kg/m^2), Green contours for mean liquid water content (kg/m^2). Each PCA value is normalised by square root of the eigen value.
Fig.3 As Fig.2, but for the last two components.
Fig.4 As Fig.2, but for sound02 channel set.
Fig.5 As Fig.4, but for the last two components.
Fig.6 Illustrating the concept of PCA scheme in the case of using two channels. Horizontal and vertical axes are brightness temperature difference for channel 1 and channel 2, respectively. Red contours mean cloud cost of clear FOV and green contours the cost of cloudy FOV. Blue lines denote threshold of PCA scheme and rectangular area surrounded by the four blue lines is declared as clear. Mean brightness temperature difference is 5K for both channels and the covariance of cloud probability is $25K^2$ for diagonal elements and $20K^2$ for off-diagonal elements, and the covariance of clear probability is $1K^2$ for diagonal elements and $0.5K^2$ for off diagonal elements.
Fig.7 Illustrating the concept of Optional PCA scheme in the case of using two channels. Horizontal and vertical axes are brightness temperature difference for channel 1 and channel 2, respectively. Red contours mean cloud cost of clear FOV and green contours the cost of cloudy FOV. Blue lines denote threshold of Optional PCA scheme and elliptical area surrounded by the blue line is declared as clear. Given statistical parameters are same as Fig.6.
Fig.8 Illustrating an example of the Optional PCA scheme for a special case. Mean brightness temperature difference is 5K for both channels and the covariance of cloud probability is $25K^2$ for diagonal elements and

24.5K^2 for off-diagonal elements, and the covariance of clear probability is 1K^2 for diagonal elements and 0.5K^2 for off diagonal elements.
Fig.9 Illustration of ECMWF scheme. Ordinal axis of upper chart means channel number sorted by cloud unaffected pressure and the abscissa means brightness temperature difference between observation minus background value. Thin black line shows original value and thick red line shows smoothed one filtered by a Low-Pass Filter. Black solid vertical line means border between clear channels and cloudy channels. Each spectral bands processed independently are devided by dashed black vertical lines. Abscissa of lower chart means cloud unaffected level derived from assuming black body cloud and the level is represented by NWP vertical model levels. Solid black line means the cloud top level corresponds to that of highest cloudy channel each spectral band.
Fig.10-1 O-B difference of cloud deteshort wave sounding channels, AIRS ch.261, ch.453, ch.672, and ch.787. Left figures are for daytime and right figures for night.
Fig.10-2 As Fig.10-1, but for AIRS ch.843, ch.914, ch.1221, and ch.1237.
Fig.10-3 As Fig.10-1, but for AIRS ch.2328, ch.2333, and AMSU-A ch.3, and ch.15.
Fig.11 O-B difference of short wave sounding channels, ch.2104 (upper), ch.2113 (middle), and ch.2333 (lower). Left figures are for daytime and right figures for night.
Fig.12 O-B difference of AMSU-A ch.2. Left figures are for daytime and right figures for night.
Fig.13-1 PCA components for cloud characterization with mix02 channel set for the daytime. The first five components and the last component are shown. Red color means positive value and violet means negative value.
Fig.13-2 As Fig.13-1, but for sound02 channel set.
Fig.14-1 As Fig.13-1, but for the nighttime.
Fig.14-2 As Fig.14-1, but for sound02 channel set.
Fig.15-1 Cloud detection map in the daytime. Blue is clear and green is cloudy. Red area means cloudy with cloud cost for Var scheme, maximum partial cloud cost for PCA, and difference of clear cost and cloudy cost for Optional PCA scheme lager than 20. The three upper rows are for Optional PCA scheme (upper), PCA scheme (middle), and Var scheme (low). Lefts are for mix02 channel set and rights for sound02 channel set. The left of the lowest row is for Mitch scheme and the right of the lowest row for ECMWF scheme.
Fig.15-2 As Fig.15-1, but for nighttime.
Fig.16-1 Cloud costs for each PCA component (PCA01-12) with mix02 channel set for all (clear and cloudy) case. Left:daytime, Right:nighttime.
Fig.16-2 As Fig.16-1, but for sound02 channel set in the nighttime.
Fig.17-1 Dependency of cloud detection to used PCA components in PCA scheme. Blue:clear (PCA01-09<2.0), Green: cloudy (PCA01-09>2.0), Yellow: cloudy (PCA01-06>2.0, PCA07-09<2.0), Red:cloudy (PCA01-06<2.0, PCA07-09>2.0). Left is for mix02 channel set and right for sound02 channel set.
Fig.17-2 As Fig.17-1, but for the nighttime.
Fig.18-1 Cloud detection map with mix02 channel set with Optional PCA scheme in the daytime. Left:with cloud probability in the nighttime, Right:with cloud probability in the daytime.
Fig.18-2 As Fig.18-1, but for sound02 channel set in the daytime.
Fig.19 Mean and standard deviation of O-B difference of clear FOV for all distributed AIRS channels and AMSU-A channels.in the daytime. Red channels are used channels in the cloud detection. The three upper rows are for Optional PCA scheme (upper), PCA scheme (middle), and Var scheme (low). Lefts are for mix02 channel set and rights for sound02 channel set. The left of the lowest row is for Mitch scheme and the right of the lowest row for ECMWF scheme.
Fig.20 As Fig.19, but for in the nighttime. The results for mix02 channel set with Optiona PCA, PCA, and Var schemes are shown.
Fig.21-1 O-B histogram for selected channels for clear case (red) and for cloudy case (green) with each schemes in the daytime. The channels are AIRS ch.261, ch.914, ch.1221, ch.2333, and AMSU-A ch.3, and ch.15 from top to bottom.
Fig.21-2 As Fig.21-1, but for the nighttime