NWP SAF

Satellite Application Facility for Numerical Weather Prediction

Document NWPSAF-EC-VS-012

Version 1.0

20 January 2006

The validation of cloud detection for AIRS

Werner Gräsle
DWD

The validation of cloud detection for AIRS

Werner Gräsle
DWD

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 16 December, 2003, 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.

NWP SAF: Visiting Scientist Report

The validation of cloud detection for AIRS

Werner Gräsle (DWD)

Host Institute:	ECMWF
Contact Point:	Tony McNally
Duration:	9 - 20 January 2006

1 Introduction

The cloud detection scheme for AIRS used by ECMWF focuses on the identification of channels not affected by clouds within each observation spot. After splitting the AIRS spectrum to 5 bands (long wave CO_2 , ozone, water vapour, 4.5μ CO_2 , and 4.2μ CO_2), the channels are ordered within any band according to their characteristic level, i.e. the level at which the relative radiance effect of an opaque black cloud to the total clear radiance exceeds 1%:

$$\frac{\left|R_{clear} - R_{cloudy}(p)\right|}{R_{clear}} > 0.01$$

where: R_{clear} = radiative flux density given a clear sky

$R_{cloudy}(p)$ = radiative flux density given an opaque black cloud with top at pressure level p

The position of a channel within the sequence ordered according to their characteristic levels is called its $rank$. No cloud contamination is assumed in channels with a characteristic level above the cloud top, and an increasing cloud impact is expected the more the characteristic level falls below the cloud top level. The first guess departure vs. channel rank curve $fg\_depar(rank)$ should display a horizontal gradient in the range of uncontaminated channels and a significantly negative or positive gradient for the channels contaminated by clouds (for cold respectively warm clouds). Therefore, the determination of the bending point of $fg\_depar(rank)$ allows the identification of the cloud contaminated AIRS channels. A low-pass filter is applied to $fg\_depar(rank)$ to reduce noise prior to the determination of the bending point. A detailed description of the method is given in McNally & Watts (2003).

A reliable exclusion of cloud contaminated radiances from the assimilation is crucial for the use of AIRS data in NWP models. Contrariwise, a cloud detection scheme tuned too much to the "safe side" would reject a huge amount of valuable, uncontaminated data, particularly in the vicinity of clouds where the availability of more data is very desirable. Therefore, this study investigates the overall performance of the ECMWF cloud detection. Additionally, data where analysed for cases of probable failure of the cloud detection (both error types) and its reasons, using the AQUA satellite's visual cloud product (MODIS).

2 Statistical analysis of the cloud detection performance

The analysis was focused on the bias corrected first guess departures fg_depar of the AIRS-787 window channel. This channel is particularly susceptible to any weakness of the cloud detection algorithm, since it is as well influenced by any cloud as it is affected in any case of unnecessary data rejection. According to the cloud detection result for AIRS-787, spots will be termed IR-clear or IR-cloudy.

A sample of 6 full days of data has been selected from the second half of 2005 (the visual cloud product cld_{vis} is not available in the data base prior to July 2005) as shown in tab.1. Unfortunately, a value of $cld_{vis} = 1$ does not always indicate clouds, it is also used for any unclarity of the visual cloud product (e. g. poor quality of visual data, sea ice, too large solar zenith angle, ...). Thus, these spots (approx. 55% of all daytime data) are not properly characterised and are of very limited value for this study.

Table 1: Characterisation of the data set used for statistical analysis.

selected analysis times	2005072500	2005072512	2005082500	2005082512
	2005092500	2005092512	2005102412	2005102500
	2005112500	2005112512	2005122500	2005122512
data excluded from statistics	- data with solar zenith angle >85° (no visual cloud product at night-time) - outer fields of view with fov<10 or fov>80 (not used for assimilation) - land and sea ice spots (not usable for window channel AIRS-787)			
amount of data	all spots	IR-cloudy	IR-clear	used in assimilation
any cld_{vis}	209464	173225	36239	9497
with $cld_{vis} < 1$	94073	62945	31128	8342

The $fg\_depar(cld_{vis})$ relationship of AIRS-787 (fig.1) offers a simple sanity check for the visual cloud product. Of cause the cloud impact should increase with cld_{vis} (i.e. the average first guess departure $\overline{fg\_depar}$ should become more negative with increasing cld_{vis}). Furthermore, the standard deviation $stdv(fg\_depar)$ should increase because the effect of variable cloud hight superimposes the impact of cld_{vis}. The observed behaviour is in good agreement with these theoretical considerations (fig.1). Particularly, the apparently perfect result for the average first guess departure $\overline{fg\_depar} = 0.001 \pm 0.008\,\mathrm{K}$ for all visually clear spots ($cld_{vis} = 0$) suggests that the visual cloud product detects virtually any cloud affecting AIRS-787. As will be shown later, an excess in the applied bias correction in the magnitude of 0.04 K relativises this "perfect" result, indicating that there might be a small amount of IR-affecting clouds not appearing visually.

As it has to be expected for an efficient cloud detection scheme, the 2D-histogram of all IR-clear spots with respect to fg_depar and cld_{vis} shows approximately Gaussian distributions roughly centred around $fg\_depar = 0$ for any cld_{vis}-bin (fig.2). The cld_{vis}-histogram of IR-clear spots (fig.3) is characterised by a predominance of visually clear spots (29%) and

spots with unknown visual cloud cover ($cld_{vis} = 1$, 14%). Nevertheless, a significant portion of all spots classified as IR-clear exhibits rather high visual cloudcover (29% have $cld_{vis} > 0.2$, 12.5% even $cld_{vis} > 0.5$). There are three possible reasons for this finding:

- There are clouds clearly detectable in the visual spectrum that do not affect the AIRS long wave CO_2 band (including AIRS-787).
- The AIRS cloud detection fails to detect a significant amount of clouds although they affect the AIRS long wave CO_2 band.
- The AQUA visual cloud product sometimes detects more clouds than present.

At least one of these reasons must apply, but also any combination of them might be possible.

In case of a perfect cloud detection scheme, the average $\overline{fg\_depar}$ of data classified as IR-clear should be zero and do not show any dependency from cld_{vis} as there is no cloud impact left in this data set. In fact the results of the ECMWF cloud detection deviate from this ideal behaviour in a statistically significant way (fig.4 and tab.2). Starting with $\overline{fg\_depar} = 0.04\,\mathrm{K}$ for $cld_{vis} = 0$, the average first guess departure becomes increasingly negative with increasing cld_{vis} until it levels off at approx. -0.10 K for $cld_{vis} > 0.2$. This clearly indicates that a noticeable amount of cloud impact passes the cloud detection undetectedly.

Table 2: Average first guess departure $\overline{fg\_depar}$ (with standard errors) and "cloud bias" CB of various groups of spots (all data in K).

		all spots	**IR-clear**	**used in assimilation**
$\boldsymbol{cld_{vis} = 0}$	$\overline{fg\_depar}$	0.001 ± 0.003	0.043 ± 0.003	-0.043 ± 0.006
	$stdv(fg\_depar)$	0.96	0.32	0.29
	CB	-0.042 ± 0.008	0 (assumed)	-0.086 ± 0.007
$\boldsymbol{cld_{vis} > 0.2}$	$\overline{fg\_depar}$	-3.25 ± 0.02	-0.099 ± 0.003	-0.153 ± 0.005
	$stdv(fg\_depar)$	3.67	0.30	0.29
	CB	-3.30 ± 0.02	-0.142 ± 0.004	-0.196 ± 0.006
any $\boldsymbol{cld_{vis}}$	$\overline{fg\_depar}$	-9.10 ± 0.03	-0.031 ± 0.002	-0.124 ± 0.003
	$stdv(fg\_depar)$	12.48	0.32	0.30
	CB	-9.15 ± 0.003	-0.074 ± 0.003	-0.167 ± 0.003

The occurrence of cloud contaminated radiances among the data classified as IR-clear necessarily causes the use of cloud contaminated data in the calculation of bias correction parameters. Consequently, the bias correction will compensate the cloud impact in the underlying data resulting in a corresponding "warming" of the data. This warming excess can be estimated as $BC_{exc} = 0.043 \pm 0.003\,\mathrm{K}$ by the observed $\overline{fg\_depar}$ of the IR-clear spots with $cld_{vis} = 0$, which presumably represent the least cloud contaminated data set available. Considering that also this data set might include some cloud contamination, the true warming ex-

cess might be even slightly larger. Taking this warming excess into account, the total average cloud impact ("cloud bias") in any data set is given by $CB_{dataset} = \overline{fg\_depar_{dataset}} - BC_{exc}$.

A "cloud bias" of $CB = -0.142 \pm 0.004$ K in the IR-clear data with $cld_{vis} > 0.2$ might be tolerable considering the accuracy of NWP models. Nevertheless, caution is demanded since it is not a "grey" bias (i.e. a bias not correlated with any meteorological variable). In fact the effect of undetected clouds will be concentrated to a range of meteorological situations being particularly prone to deceive the cloud detection scheme. Those situations will systematically be correlated with a "cloud bias" significantly more negative than -0.14 K. Furthermore, one would expect these situations to appear more often in areas with a heterogeneous patchy cloud distribution than in extended almost cloudless areas. As a consequence, spots wrongly classified as IR-clear will "prefer" areas where only few spots are IR-clear and hence will have a better chance to survive the thinning preceding the data assimilation into the NWP model than spots correctly classified as IR-clear. This statistical preference of spots wrongly classified as IR-clear by the thinning process is confirmed by the statistical analysis showing a larger $CB = -0.196 \pm 0.006$ K for $cld_{vis} > 0.2$ (which represent those situations with patchy cloud distribution) in the used data than in all IR-clear data (fig.4 and tab.2).

Next, an attempt will be made to determine the normalised histogram $p_{wc}(fg\_depar)$ of fg_depar for the spots wrongly classified as IR-clear, and the fraction F_{wc} of cloud contaminated data among the IR-clear data resp. among the used data. The mathematical derivation is a bit complicated and can be skipped without compromising the understanding of the results.

Beforehand, $p_{wc}(fg\_depar)$ is not known because the spots wrongly classified as IR-clear cannot be identified. However, under certain conditions a reasonable estimation of this distribution is possible. Assuming that the histogram $p_{tc}(fg\_depar)$ of spots correctly classified as IR-clear is symmetrical with respect to $fg\_depar = BC_{exc}$, $p_{wc}(fg\_depar)$ can be approximated by the asymmetric part of the histogram $p_c(fg\_depar)$ of all IR-clear spots. Obviously, this approximation will be poor if $p_{wc}(fg\_depar)$ is not essentially separated from $fg\_depar = BC_{exc}$, i.e. if a significant part of $p_{wc}(fg\_depar)$ extends to fg_depar values greater than BC_{exc}. Fig.5 shows the relevant histograms both on the basis of a perfect bias correction $BC_{exc} = 0$ and assuming a more realistic (but possibly still to optimistic) warming excess of $BC_{exc} = 0.04$ K. The resulting fractions F_{wc} of cloud contaminated data in the IR-clear respectively in the used data as well as their average first guess departures $\overline{fg\_depar_{wc}}$ are given in tab.3.

In case that the true $p_c(fg\_depar)$ is not essentially confined to $fg\_depar < BC_{exc}$, the true $p_{wc}(fg\_depar)$ would be somehow intermediate between $p_c(fg\_depar)$ and the calculated $p_{wc}(fg\_depar)$. Therefore, the calculated F_{wc}-values must be understood as lower limits, and the true $\overline{fg\_depar_{wc}}$ might be less negative than calculated.

The derived "cloud bias" for spots wrongly classified as IR-clear in the range of -0.38 to -0.43 K might not be considered negligible.

Table 3: Calculated average first guess departures $\overline{fg\_depar_{wc}}$, "cloud bias" CB and frequency F_{wc} of spots with $cld_{vis} > 0.2$ wrongly classified as IR-clear (all data in K).

$cld_{vis} > 0.2$	BC_{exc} = 0		BC_{exc} = 0.04 K	
	IR-clear	**used in assimilation**	**IR-clear**	**used in assimilation**
$\overline{fg\_depar_{wc}}$	-0.404 stdv = 0.18	-0.382 stdv = 0.17	-0.386 stdv = 0.18	-0.365 stdv = 0.18
CB	-0.404	-0.382	-0.426	-0.405
F_{wc} [%]	24.7	38.1	34.7	48.1

The fraction F_{wc} of cloud contaminated data differs significantly between all IR-clear spots and used spots. During the thinning process the fraction of spots with $cld_{vis} > 0.2$ wrongly classified as IR-clear increases from 25% to 38% (assuming $BC_{exc} = 0$) or from 35% to 48% (for $BC_{exc} = 0.04\,\mathrm{K}$). In contrast, the histograms $p_{wc}(fg\_depar)$ are very similar (with respect to centre and width of the distribution) for IR-clear spots and used spots (tab.3). This finding corresponds to the expectation, that the thinning algorithm indeed favours cloud contaminated spots because of their spatial distribution, but does not cause any systematic discrimination within the population of spots wrongly classified as IR-clear.

These findings apply to visual cloudy spots ($cld_{vis} > 0.2$). To estimate the frequency F_{wc} of spots wrongly classified as IR-clear among all IR-clear respectively all used observations, the average "cloud bias" CB_{wc} in a set of spots wrongly classified as IR-clear is assumed to be independent from the average visual cloudcover of the set. As shown for visual cloudy spots a typical value is $CB_{wc} = -0.4\,\mathrm{K}$.

Applying $\overline{fg\_depar_{wc}} = CB_{wc} + BC_{exc}$

and $\overline{fg\_depar_{tc}} = BC_{exc}$

and $F_{tc} + F_{wc} = 1$

to $\overline{fg\_depar_{c}} = F_{tc}\ \overline{fg\_depar_{tc}} + F_{wc}\ \overline{fg\_depar_{wc}}$

we yield $\overline{fg\_depar_{c}} = BC_{exc} + F_{wc}\ CB_{wc}$

This results in $$F_{wc} = \frac{\overline{fg\_depar_{c}} - BC_{exc}}{CB_{wc}}$$

According to this estimation and assuming a perfect bias correction ($BC_{exc} = 0$), about 7.7% of all IR-clear spots are cloud contaminated (tab.4). The selectivity effect of the thinning procedure is quite dramatic rising this fraction to 31%. Assuming a more likely $BC_{exc} = 0.04\,\mathrm{K}$,

about 18% of all IR-clear spots are affected by clouds. **After thinning we end up with 41% cloud contaminated observations in the assimilated data, corresponding to an average "cloud bias" of $CB_{used} = -0.164 \pm 0.004$ K .**

Table 4: Estimated fraction F_{wc} of observations wrongly classified IR-clear and "cloud bias" CB_c in all IR-clear respectively used spots. The estimation of F_{wc} is based on an average $CB = -0.4$ K in all spots wrongly classified IR-clear.

any cld_{vis}	BC_{exc} = 0		BC_{exc} = 0.04 K	
	IR-clear	**used in assimilation**	**IR-clear**	**used in assimilation**
F_{wc} [%]	7.7	30.9	17.7	40.9
CB_c [K]	-0.031 ± 0.002	-0.124 ± 0.003	-0.071 ± 0.004	-0.164 ± 0.004

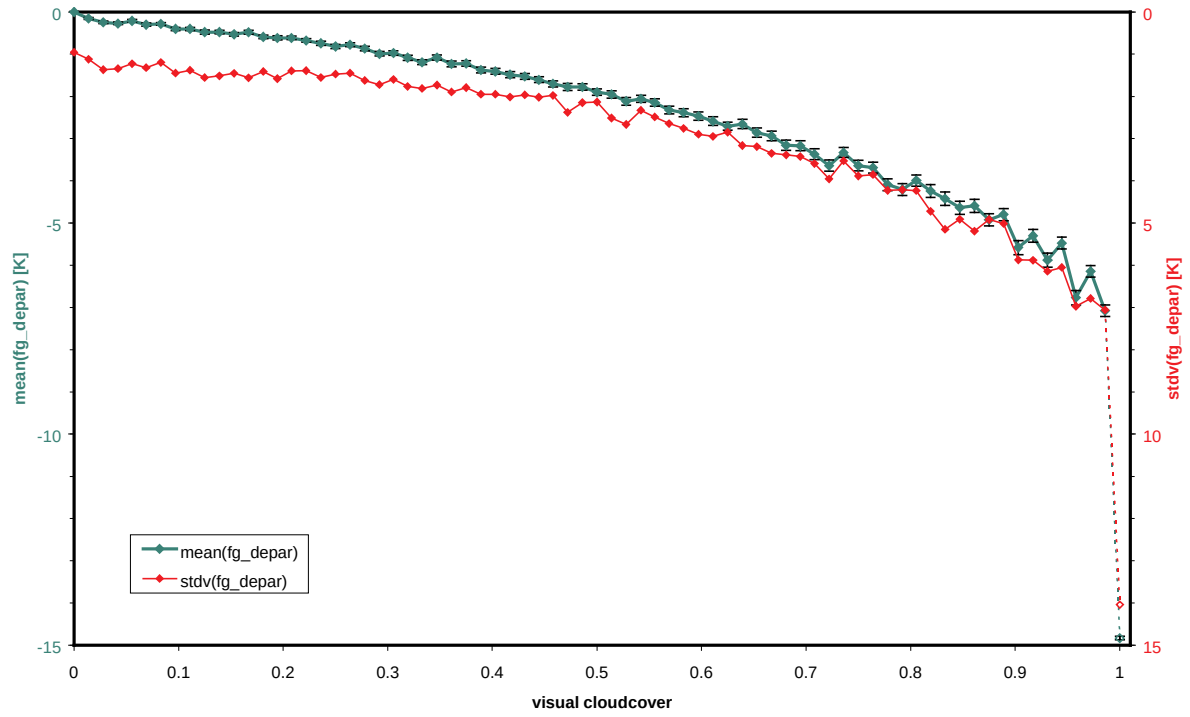

Figure 1: The $fg\_depar(cld_{vis})$ relationship for all spots in the statistical analysis.

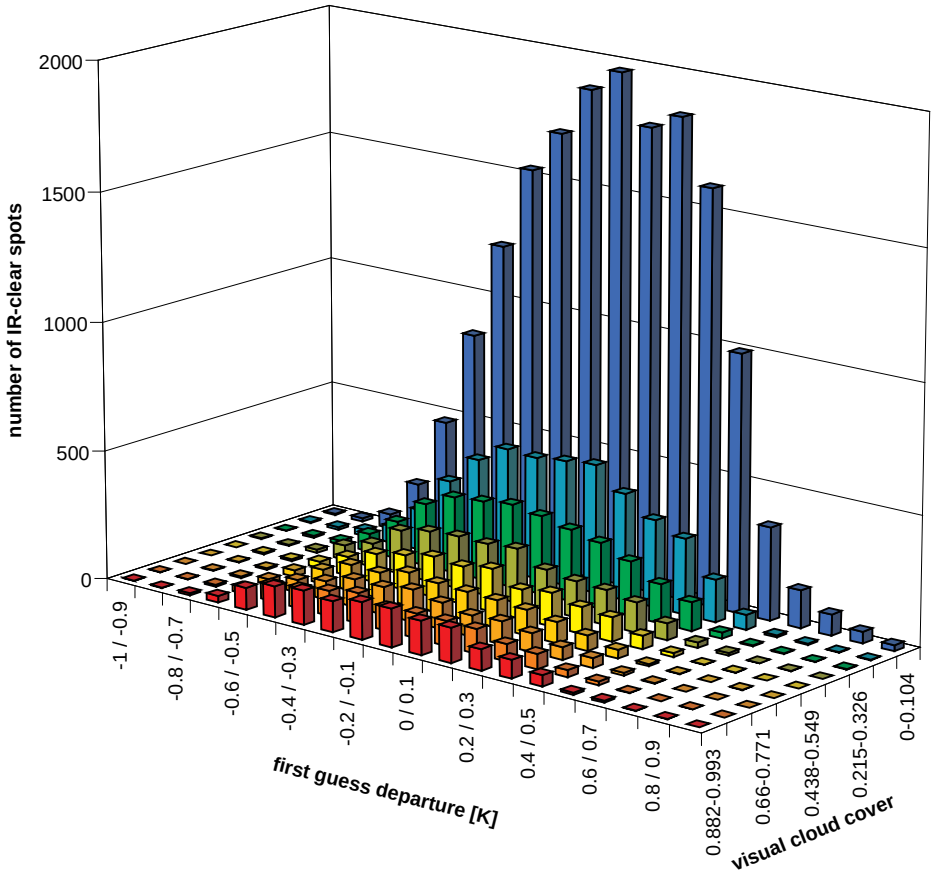

Figure 2: 2D-histogram of all IR-clear spots with respect to first guess departure fg_depar and visual cloud-cover cld_{vis}.

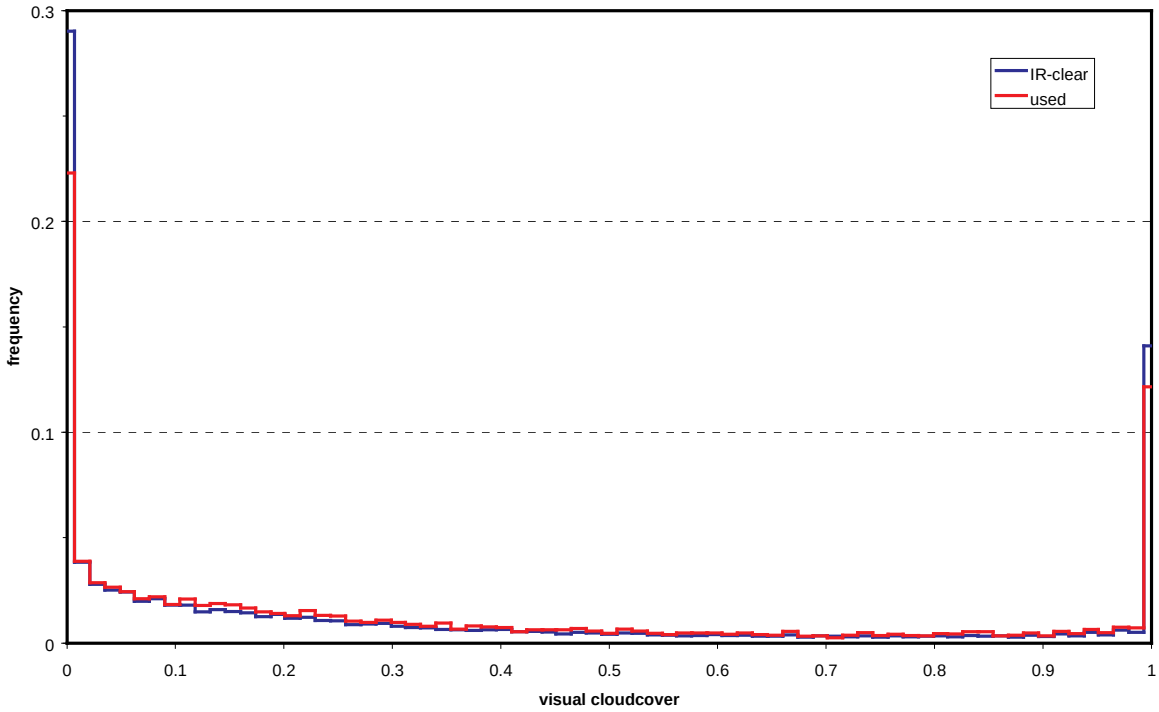

Figure 3: Normalised histograms of IR-clear and used spots with respect to visual cloudcover.

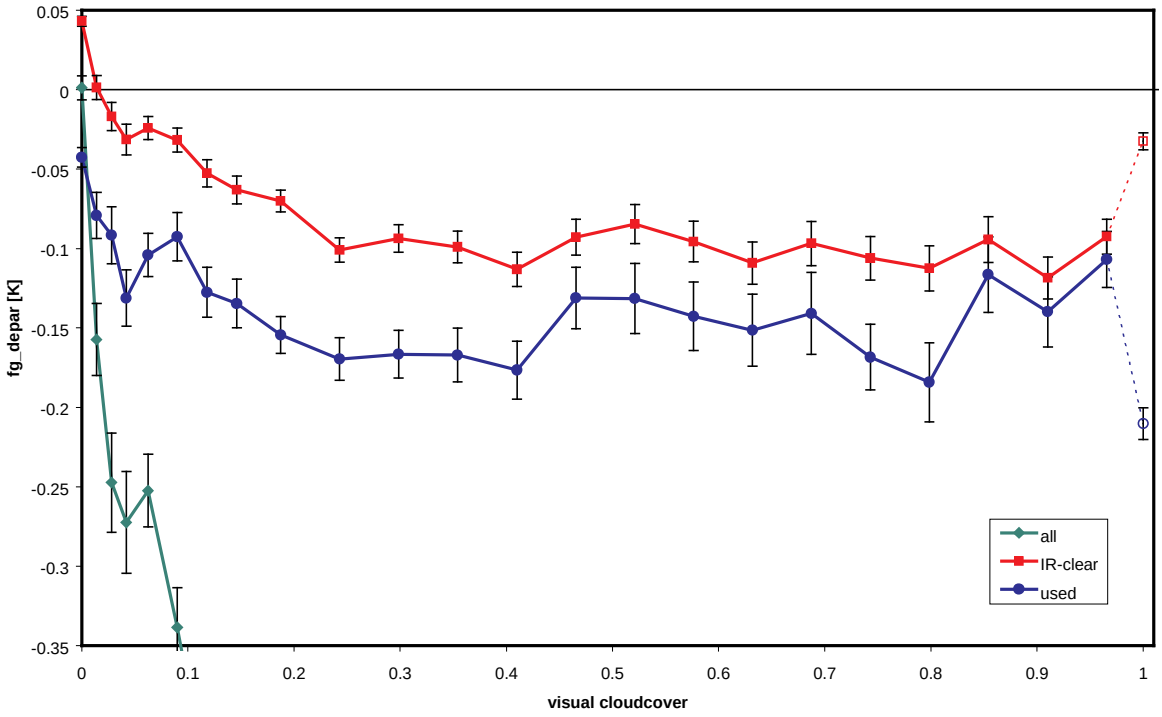

Figure 4: cld_{vis}-dependent fg_depar for all, IR-clear, and used spots (note, that the values for $cld_{vis}=1$ include many spots with an unknown visual cloud cover and can therefore not be interpreted as fully cloudy).

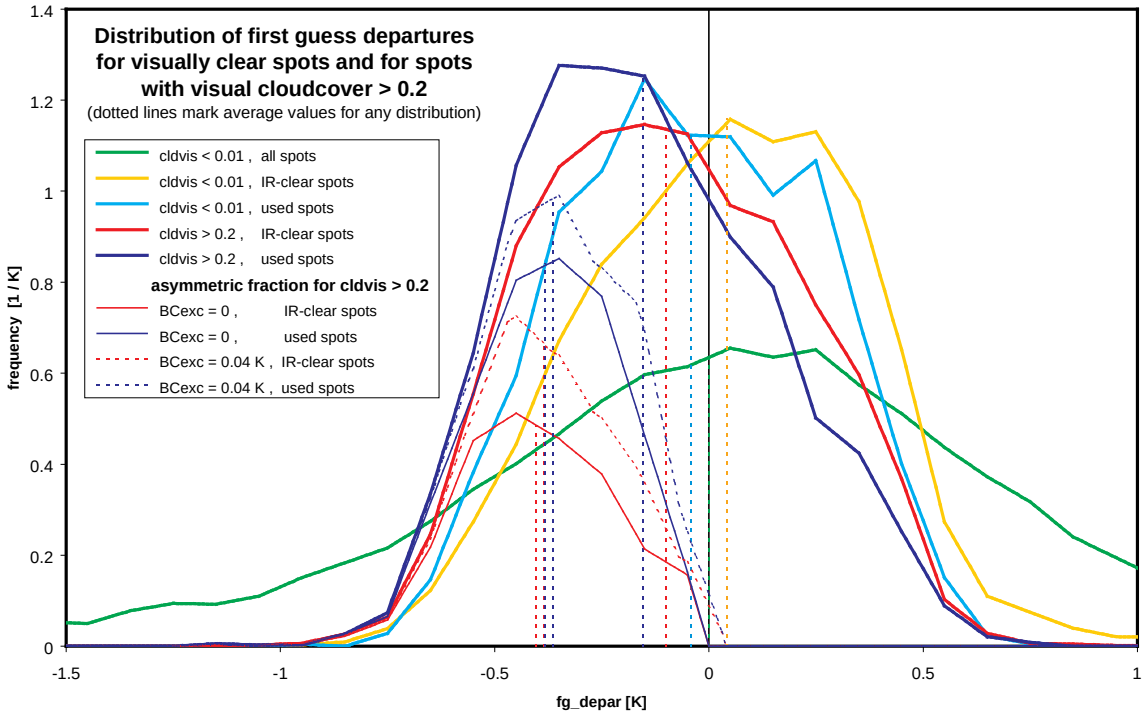

Figure 5: fg_depar -distributions of IR-clear as well as used data for visually clear ($cld_{vis} = 0$) and visually significantly cloudy ($cld_{vis} > 0.2$) spots. The asymmetric fraction (calculated for both a perfect bias correction and a warming excess of $BC_{exc} = 0.04\,\mathrm{K}$) of the latter distributions are approximations for the distribution $p_{wc}(fg\_depar)$ of spots wrongly classified as IR-clear.

3 Case 1: IR -clear spots with high visual cloudcover

As shown by the statistical analysis, the probability of the cloud detection to miss a cloud (type 1 error) is correlated with the visual cloudcover (chap.2). The best chance to find cases of type 1 errors would be checking spatially isolated IR-clear spots with high cld_{vis}. However, analysing a sample of this type has two major disadvantages:

- Developing a filter for extracting this sample would be rather complicated (since only IR-clear spots isolated because of the meteorological situation are of interest, the filter has to distinguish this situation from isolation caused by the land-sea-distribution)
- This sample would be very heterogeneous with respect to the meteorological situation and the possible reason of the cloud detection failure. This probably might hamper or even prevent the identification of the reasons of type 1 errors.

Therefore, another approach was chosen. The data were searched for a situation, where a considerable number of IR-clear spots exhibiting high cld_{vis} can be found in a meteorologically rather concise situation. An appropriate situation was found for analysis time 20051225 00UTC in a region area east of Hawaii (14° < lat < 25°, -175° < lon < -155°). As shown in fig.6 there are numerous IR-clear spots (large green circles) with high cld_{vis} (yellow, red or purple dots) particularly in the southern part of the area covered by model clouds.

Obviously, the cloud detection fails in a sense, that it misses many spots with high cld_{vis} even up to full visual cloud coverage. This does not necessarily mean that it fails with respect to its purpose, i. e. that it fails to identify IR-radiances affected by clouds. A plot of fg_depar vs. cld_{vis} for all 232 IR-clear spots with known cld_{vis} in the investigated area (fig.7) shows virtually no dependency ($d\,fg\_depar/d\,cld_{vis} = -0.02$ K) and a $\overline{fg\_depar}$ of 0.076 ± 0.017 K (corresponding to a very small "cloud bias" of $CB = 0.03 \pm 0.02$ K). Thus, there is no indication of any cold cloud impact passing the cloud detection in this sample.

Essentially, there are two scenarios explaining the occurrence of high visual cloudcover without any detectable impact on the AIRS-787 window channel:

- A warm low cloud or fog cover with a cloud top temperature almost perfectly resembling the sea surface emission.
- A high thin ice cloud (cirrus) strongly interacting with radiation in the visual spectrum but almost completely transparent for AIRS-787.

An analysis of the NWP models atmospheric temperature and humidity profiles at the locations of the IR-clear spots was carried out to decide this alternative. Average profiles for six sub-samples characterised by different ranges of cld_{vis} are shown in fig.8. The overall shape of the profiles is rather similar in all cld_{vis}-classes with distinctive maxima of relative humidity $RH(p)$ around 100 hPa as well as from 900 hPa down to the surface, thus excluding neither of the two scenarios. When focusing on the deviations of these average profiles from the average profiles of the visual clear spots ($cld_{vis} = 0$), some significant relationships between

atmospheric state variables and observed cld_{vis} can be recognised (fig.9). Particularly, a high $RH(p)$ coinciding with a positive correlation of cld_{vis} and $RH(p)$ would be an indication of the possible occurrence of clouds at that pressure level.

This situation is clearly given at 100 hPa supporting the assumption that cirrus clouds account for the observed cld_{vis}. But there is also an albeit small increase of $RH(p)$ with cld_{vis} at the surface together with $RH(1000\ \text{hPa}) \approx 80\%$. Thus the occurrence of a fog layer as a possible reason of the observed cld_{vis} can not be rejected definitely, although it appears to be less probable than the cirrus layer. Further evidence for the cirrus hypothesis arises from the fact that two other indicators of a relationship between $RH(p)$ and cld_{vis}, namely the slope $d\,RH(p)/d\,cld_{vis}$ and the coefficient of correlation, both show a much more distinct maximum at 100 hPa than at 1000 hPa (fig.10).

Moreover, the NWP model shows lots of high clouds but almost no low clouds in the central part of the investigated area where many IR-clear spots coincide with high cld_{vis} values (fig.11+12). There are no visual clouds whatsoever observed in the northern part of the investigated area (north of 20°N) where high level model clouds occur. Thus, the high cloud diagnostic of the model seems to be unreliable in this particular situation.

Another general meteorological consideration also favours the cirrus hypothesis: Fog usually develops in situations characterised by a significant temperature difference between the surface and the overlying air mass. Therefore, the formation of a homogeneous fog layer on a scale of several hundreds of kilometres does not show any significant radiative signature in the long wave CO_2 band is a very unlikely phenomenon.

Altogether, high cirrus clouds around 100 hPa without much doubt account for the observed visual cloudcover in IR-clear spots. It perfectly goes with the purpose of the AIRS cloud detection scheme to "miss" these cirrus clouds, because they obviously do not affect the AIRS long wave CO_2 band and therefore cause no problems in the assimilation to NWP models.

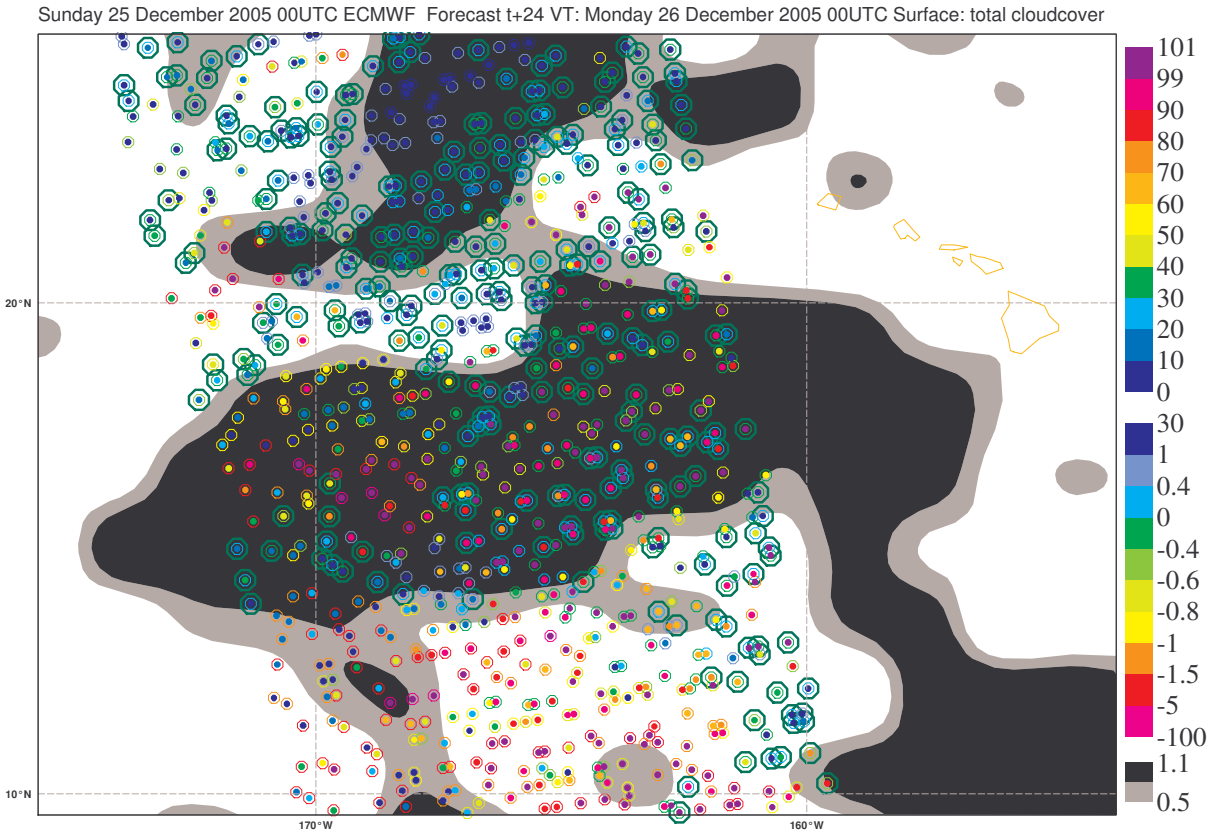

Figure 6: Satellite observations, total model clouds and cloud detection results in the investigation area east of Hawaii. Total model cloudcover is shown by grey (>0.5) and black (>0.7) shading. Visual cloudcover is represented by the dots (upper colour scale, in %), whereas the colour of the small surrounding circles indicates AIRS-787 first guess departures (lower colour scale). IR-clear spots are indicated by large green circles.

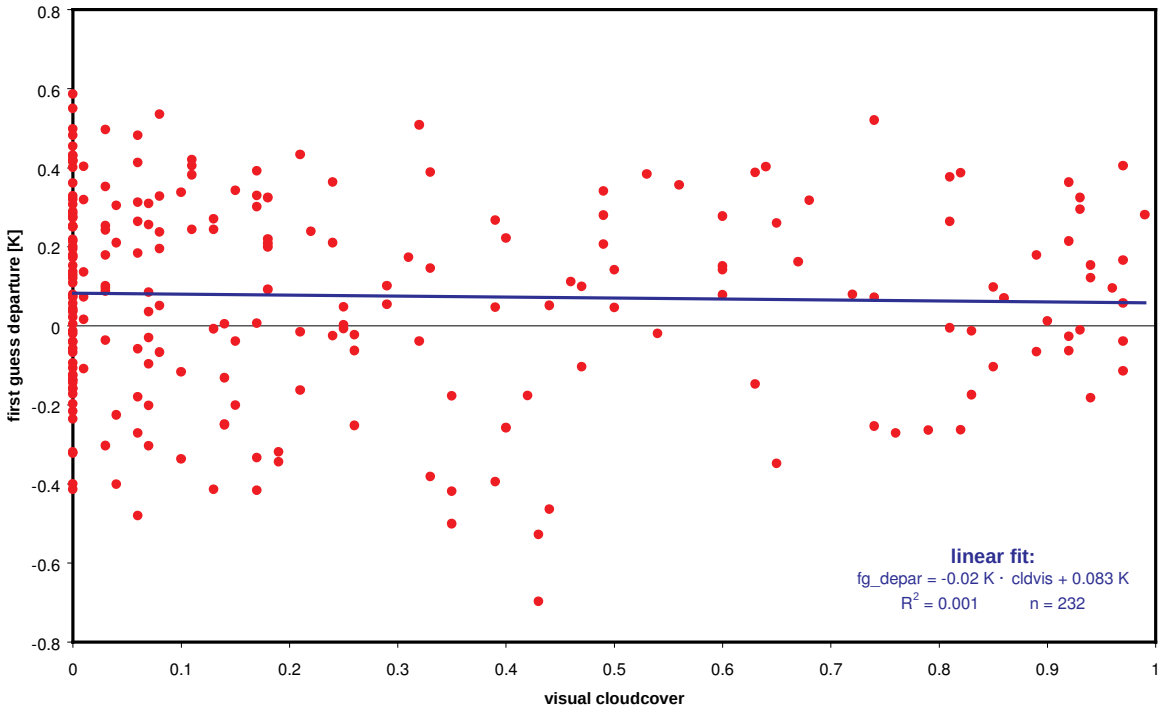

Figure 7: No correlation between first guess departure fg_depar and visual cloudcover cld_{vis} can be found among the IR-clear spots with known cld_{vis} in the investigation area east of Hawaii.

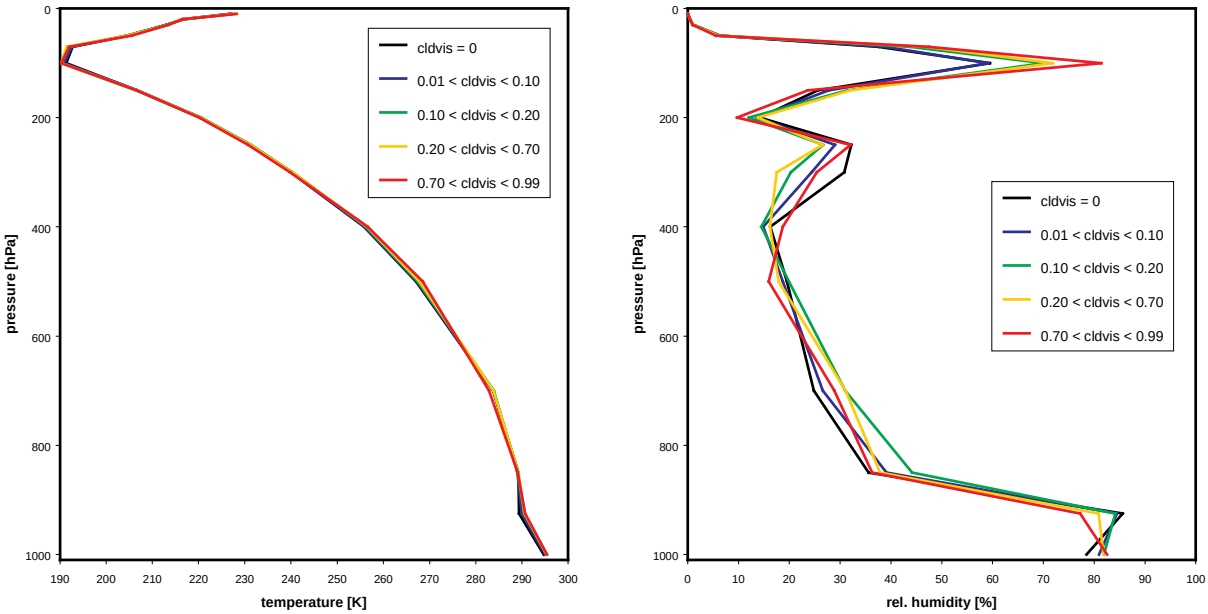

Figure 8: Average model temperature and relative humidity profiles for several cld_{vis}-classes of IR-clear spots.

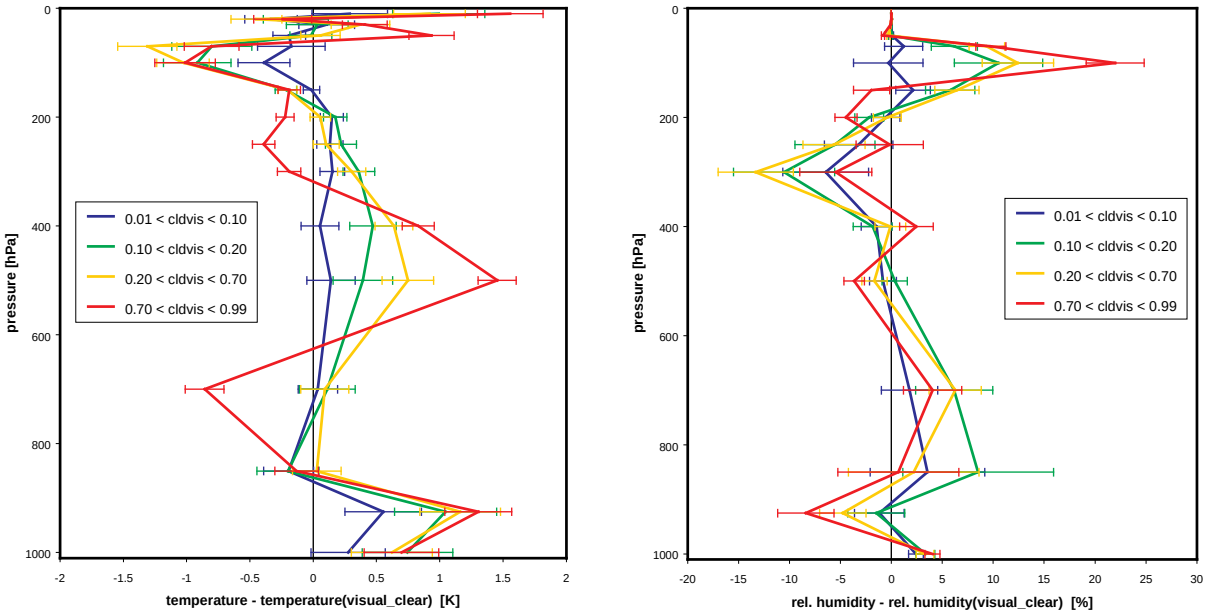

Figure 9: Average differences of model temperature and relative humidity profiles for several cld_{vis}-classes of IR-clear spots relative to the average profiles of visual clear spots ($cld_{vis} = 0$).

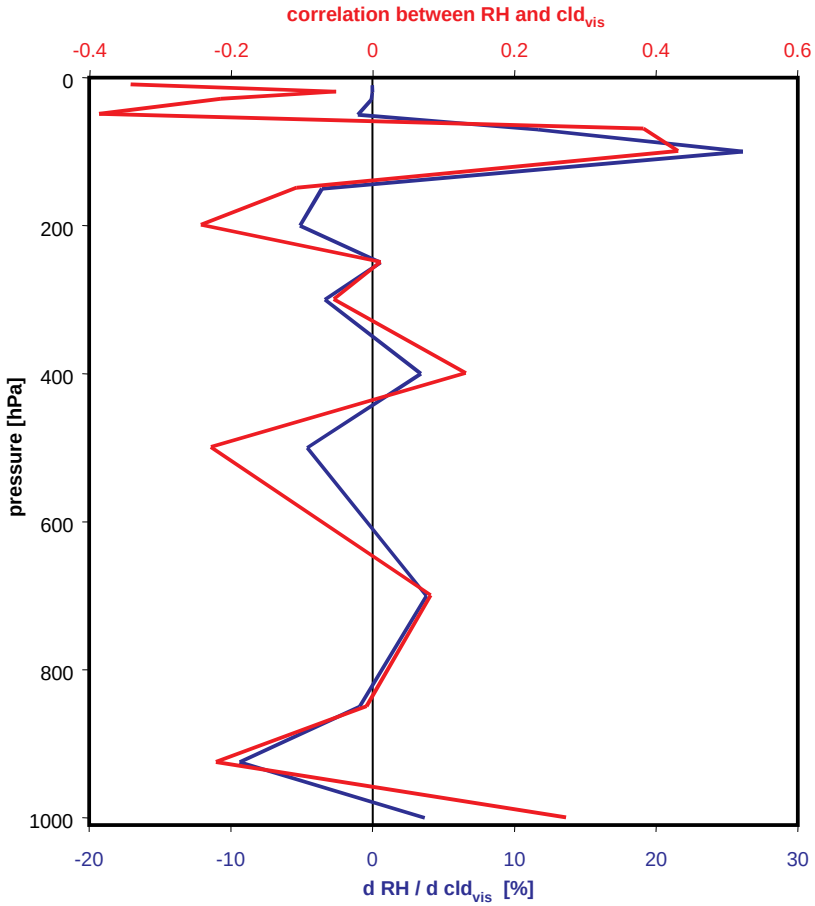

Figure 10: The coefficient of correlation and slope of the $RH(p)$ to cld_{vis} relationship of IR-clear spots.

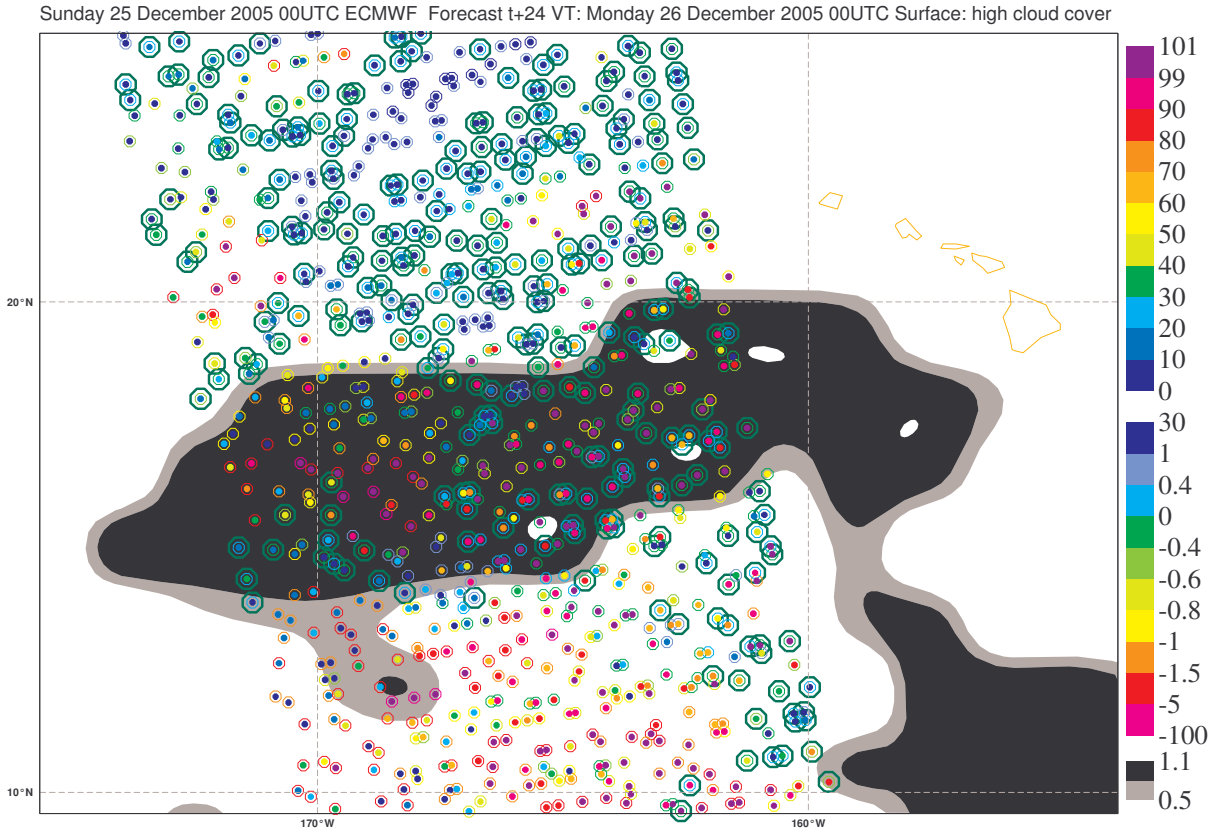

Figure 11: Satellite observations, high level model clouds and cloud detection results in the investigation area east of Hawaii. For the legend see fig.6.

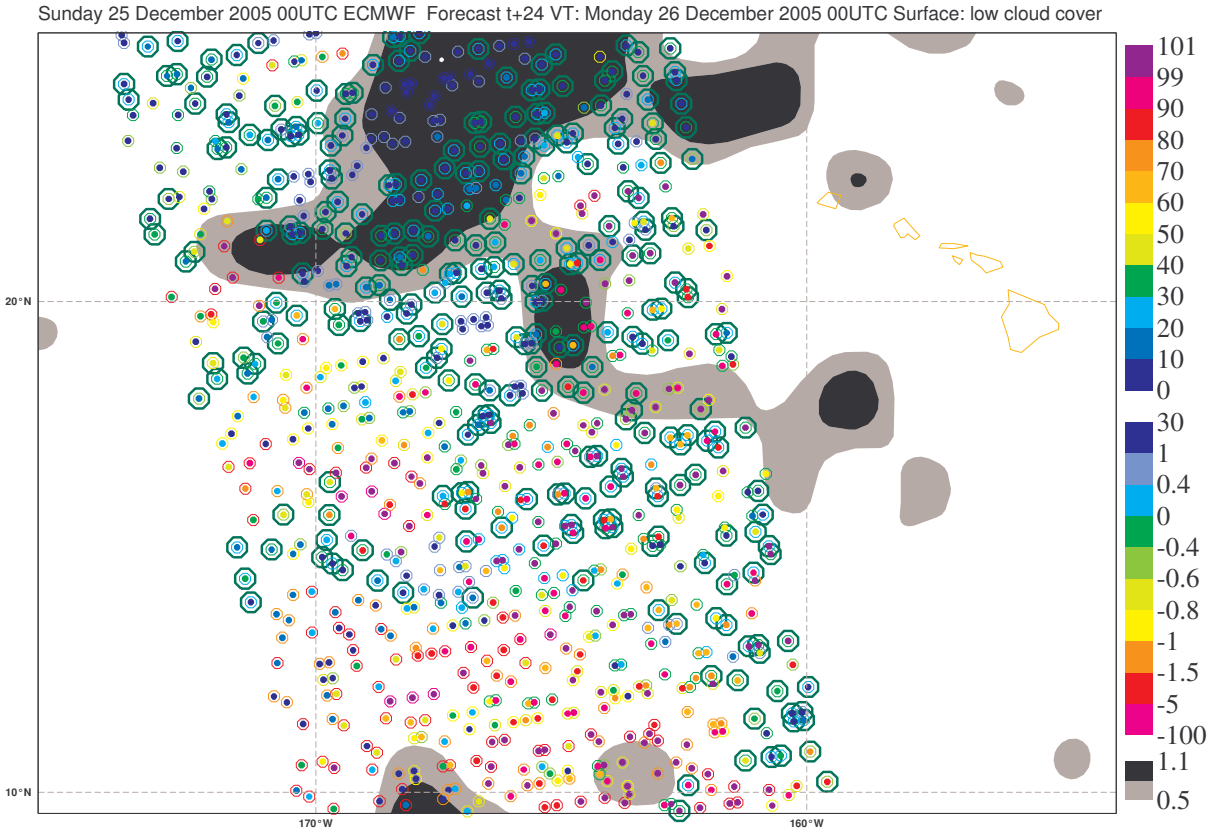

Figure 12: Satellite observations, low level model clouds and cloud detection results in the investigation area east of Hawaii. For the legend see fig.6.

4 Case 2: IR -cloudy spots without any visual cloudcover

Apart from missing clouds, rejecting too many truly clear data (type 2 error) is another common problem in cloud detection. The chosen region (fig.13) south of Baja California (analysis time 20051225 00UTC, $8.5° < \text{lat} < 19°$, $-119° < \text{lon} < -110°$) appears appropriate to investigate this issue, because there are many IR-cloudy observations in its northern part although the visual cloud product indicates a fully clear situation ($cld_{vis} = 0$, blue dots in fig.13). The southern part of the area is covered by a strong tropical convective cloud cluster causing a fg_depar up to -65 K. The average relative humidity profiles $RH(p)$ from the NWP model for various classes of observations (fig.14) exhibit the characteristic deep convection signature in the heavily clouded part with distinctively increased $RH(p)$ throughout the profile up to approximately 150 hPa.

Corresponding to the subject of type 2 errors, the following considerations refer to the visual clear spots only ($cld_{vis} = 0$). Since $RH_{cld_{vis}=0}(p) < 50\%$ applies anywhere above 900 hPa in the $RH(p)$-profiles (fig.14), any IR-affecting cloud contamination possibly occurring in visual clear spots must be located close to the surface. Only minor differences in $RH(p)$ can be found between IR-cloudy and IR-clear spots. Nevertheless, the small but statistically highly significant increase in the average relative humidity of IR-cloudy observations at the lowest pressure level $RH_{IR-cloudy}(1000\,\text{hPa}) - RH_{IR-clear}(1000\,\text{hPa}) = 4.7 \pm 1.2\%$ (fig.14) deserves attention. An average $RH_{IR-cloudy}(1000\,\text{hPa}) = 84.9\%$ might indicate a considerably higher probability of near-surface cloud formation in the IR-cloudy spots than in the IR-clear spots which have $RH_{IR-clear}(1000\,\text{hPa}) = 80.2\%$.

The differences of average model temperature profiles between the IR-cloudy and IR-clear spots (fig.15) exhibit significantly cooler conditions in the IR-cloudy spots below 900 hPa reaching a maximum difference of $T_{IR-cloudy}(1000\,\text{hPa}) - T_{IR-clear}(1000\,\text{hPa}) = -1.16 \pm 0.56\,\text{K}$. In contrast the model sea surface temperature T_{surf} is very homogeneous and exhibits virtually no difference between IR-cloudy and IR-clear spots. Thus, the IR-cloudy and IR-clear spots represent groups of significantly different atmospheric model conditions.

To figure out whether the IR-cloudy spots are examples of type 2 errors, it has to be checked if there are variations in the AIRS observations corresponding to the differences in model conditions. If not, the IR-cloudy classifications are most likely type 2 errors caused by first guess errors (i. e. model errors). Otherwise a true cloud contamination in the IR-cloudy spots is quite probable (but not proven).

As shown in fig.16 the average measured brightness temperatures bt_obs for used mid- and lower-tropospheric channels ($20 \le rank \le 54$) in the long wave CO_2-band of IR-cloudy spots are significantly warmer than those of the IR-clear spots (typically $bt\_obs_{IR-cloudy} - bt\_obs_{IR-clear} = 0.76 \pm 0.58\,\text{K}$). For the sake of clarity the channels are ordered in fig.16 according to the ranking scheme explained in the introduction. The calculated first guess differences $fg_{IR-cloudy} - fg_{IR-clear}$ resemble the behaviour of the bt_obs-

differences. $fg_{IR-cloudy} - fg_{IR-clear}$ even exceeds $bt\_obs_{IR-cloudy} - bt\_obs_{IR-clear}$ in the lower troposphere down to the surface ($rank > 33$). Thus, IR-clear and IR-cloudy spots obviously form physically different samples, and the rejection of the IR-cloudy spots might be well-founded.

To recognise why the AIRS cloud detection classifies more than 85% of all visual clear spots in the investigation area as IR-cloudy for the AIRS-787 window channel, the fg_depar of all used tropospheric channels in the long wave CO_2 band have been studied. The AIRS cloud detection uses a running average $av21\_fg\_depar$ of the first guess departures in the *rank*-ordered space, averaging over 21 observations within the long wave CO_2 band. Averaging the $av21\_fg\_depar$-curves of all IR-clear spots results in a curve which is rather flat up to $rank = 50$ and increases with a slope of 0.017 K per rank beyond (fig.17). The IR-clear spots form a rather homogeneous group, resembling the shape of the average curve within any single curve (fig.18). None of the individual slopes exceeds 0.025 K per rank, thus being below the chosen slope-threshold for cloud detection.

The average $av21\_fg\_depar$-curve of the IR-cloudy spots differs from that of the IR-clear ones by descending to a distinct minimum before bending into a similar positive slope around $rank = 50$ (fig.17). Furthermore, the IR-cloudy spots constitute a much more heterogeneous group (fig.19) with a standard deviation up to 0.53 K for $rank > 50$ compared with 0.12 K for the IR-clear spots. Despite their minor average near-surface slope of 0.016 K per rank, the slopes for most individual spots clearly exceeds 0.025 K per rank (resp. undershoots -0.025 K per rank in the descending part of the curve). Those spots that have only minor slopes are classified as IR-cloudy because they exhibit a distinct minimum.

Altogether, although the IR-clear spots are very similar to some of the IR-cloudy spot (cf. fig.18 to some of the curves in the 3rd and 4th plot of fig.19), the IR-cloudy spots form a much more heterogeneous group that exhibits significant distortions attributable to either cloud impact or errors in the model temperature profiles (including sea surface temperature T_{surf}).

There are some indications of model errors affecting the cloud detection in the investigated case, particularly of errors in the model sea surface temperature T_{surf}. Almost all $av21\_fg\_depar$-curves of visual clear spots exhibit a significant increase for $rank > 50$, regardless whether they are IR-clear or IR-cloudy (fig.18+19). A negative bias of the model T_{surf} in the investigation area provides a self-evident explanation for this behaviour. The assumption that too low near-surface temperatures in the model amplify this effect for some of the IR-cloudy spots (particularly those in the 4th plot of fig.19) corresponds to the observed negative $T_{IR-cloudy} - T_{IR-clear}$-differences below 900 hPa.

The $av21\_fg\_depar$-curves of many visual cloudy spots characterised by a decrease down to a minimum near $rank = 50$ before bending into an increasing near-surface branch can then be caused in two different ways (or a combination of both):

- The impact of a cold cloud causing the descending branch, which is partly compensated in the surface-sensitive channels by the T_{surf} -bias.
- Too warm model conditions in the mid-troposphere would yield a similar result. In this case, the lower sensitivity of the window channel to the mid-tropospheric temperature would also contribute to the increase of $av21\_fg\_depar$ at high $rank$. The positive $T_{IR-cloudy} - T_{IR-clear}$ -differences between 500 and 700 hPa might point at this possibility. But the correlations between fg_depar and $T(500\,\mathrm{hPa})$ for IR-cloudy spots are very weak ($r^2 < 0.1$ for any channel, even those around $rank = 50$). Thus, there is no striking support for this explanation.

Consequently, a true cloud impact in the IR-cloudy observations cannot be excluded, although there is strong evidence that errors in the model temperature profiles and sea surface temperature play an important role for the observed fg_depar -distortions.

Therefore, the cloud detection scheme does a good job rejecting many of the visual clear spots in the investigation area. As long as there is no robust criterion available to distinguish, whether the distortion of a $av21\_fg\_depar$ -curve is caused by cloud impact or results from model errors, flagging all these spots as IR-cloudy is the only passable treatment. Nevertheless, as the assimilation of radiances currently rejected because of errors in the model temperature profiles would be very desirable to reduce these model errors, pushing the development of such a robust criterion is highly recommended.

Finally, another possible problem capable to cause type 2 errors has been recognised. The visual clear observations spot 30902 (IR-clear) and 31940 (IR-cloudy) are located within a distance of less than 20 km from each other (fig.13). Since the slopes occurring in their $av21\_fg\_depar$ -curves (fig.20) are below the threshold for cloud detection, the classification of spot 31940 as IR-cloudy is based on the existence of a distinct minimum in its $av21\_fg\_depar$ -curve. The curve of spot 30902 also exhibits a minimum, but it is less pronounced and its depth is below the corresponding threshold for cloud detection.

An eye-catching difference between the fg_depar of both spots is the existence of some distinct outliers within the high-$rank$ range ($rank = 47, 57, 62$) for spot 31940 (fig.20). Of cause these outliers affect the shape of the $av21\_fg\_depar$ -curve, particularly AIRS-300 ($rank = 47$) contributes significantly to the depth of the minimum as it is shown in fig.21. Even if this might not be decisive for the classification of the spot as IR-cloudy in this case, it nevertheless highlights that outliers have the potential to affect the cloud detection in a relevant way. Since outliers almost definitely result from various types of measurement distortions and thus do not represent the state of the atmosphere[1)], it is desirable to reduce the impact of outliers on the cloud detection as far as possible. The use of a well-maintained black-

[1)] A state of the atmosphere that yields very different fg_depar for AIRS channels with very similar sensitivity characteristics (as those adjacent in *rank*-ordered space) would almost certainly be physically unrealistic.

listing of suspicious channels and the utilisation of the channel quality flag distributed with any single AIRS observation are important instruments in this effort.

Another outlier-filter has been developed. First the running average $av9\_fg\_depar$ and the corresponding standard deviation $stdv9\_fg\_depar$ spanning 9 subsequent channels in $rank$ -ordered space are calculated. Then all channels fulfilling both

$$\left| fg\_depar(rank) - av9\_fg\_depar(rank) \right| > \Delta_{outlier}^{abs}$$

and

$$\left| fg\_depar(rank) - av9\_fg\_depar(rank) \right| > \Delta_{outlier}^{rel} \cdot stdv9\_fg\_depar(rank)$$

are classified as outliers and are eliminated from the data set used for cloud detection and assimilation. The impact of this means of quality control for both spots is shown in fig.22. With thresholds of $\Delta_{outlier}^{abs} = 0.6\,\mathrm{K}$ and $\Delta_{outlier}^{rel} = 1.5$ the number of excluded outliers is 5 out of 71 for spot 30902 respectively 6 for spot 31940. The yield in robustness by this method is evident, and the smoothing effect on the $av21\_fg\_depar$ -curves looks promising. Further investigation of its practical benefit within a data assimilation system as well as a tuning of $\Delta_{outlier}^{abs}$, $\Delta_{outlier}^{rel}$, and the width of the running average is required.

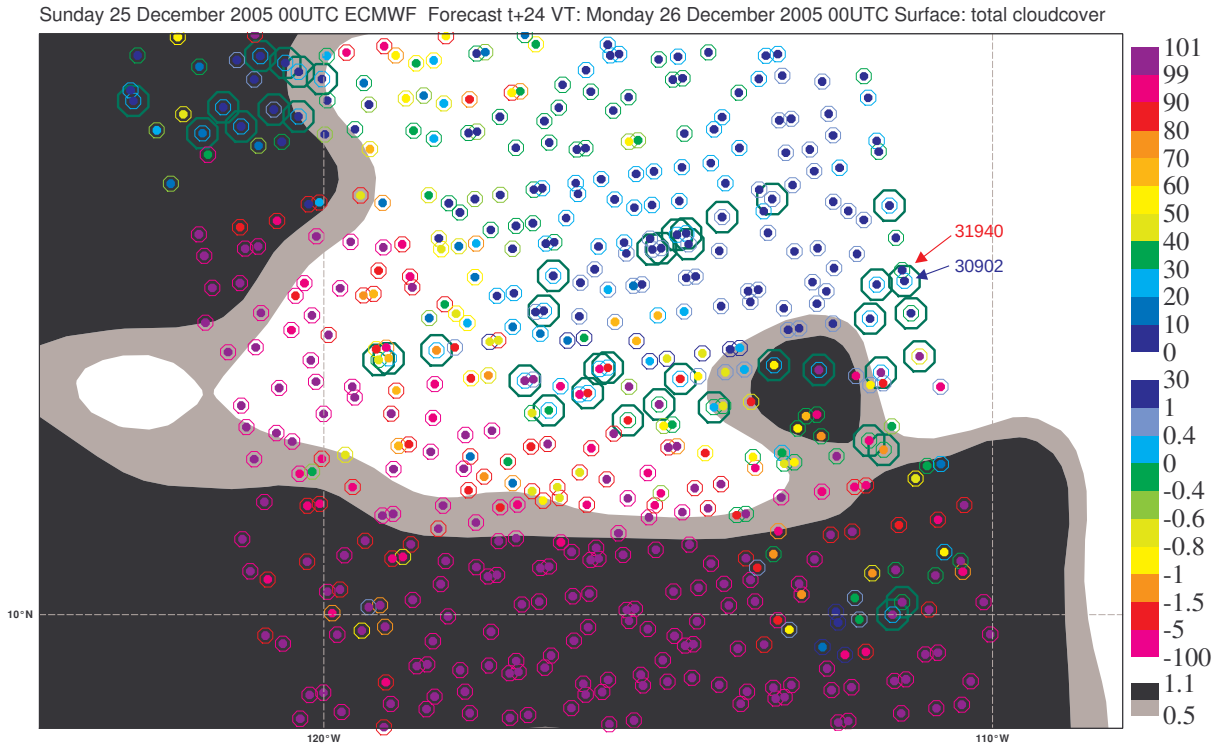

Figure 13: Satellite observations, total model clouds and cloud detection results in the investigation area south of Baja California. Total model cloudcover is shown by grey (>0.5) and black (>0.7) shading. Visual cloudcover is represented by the dots (upper colour scale, in %), whereas the colour of the small surrounding circles indicates AIRS-787 first guess departures (lower colour scale). IR-clear spots are indicated by large green circles. Note that the heterogeneous situation west of -119° longitude is not included in the investigation.

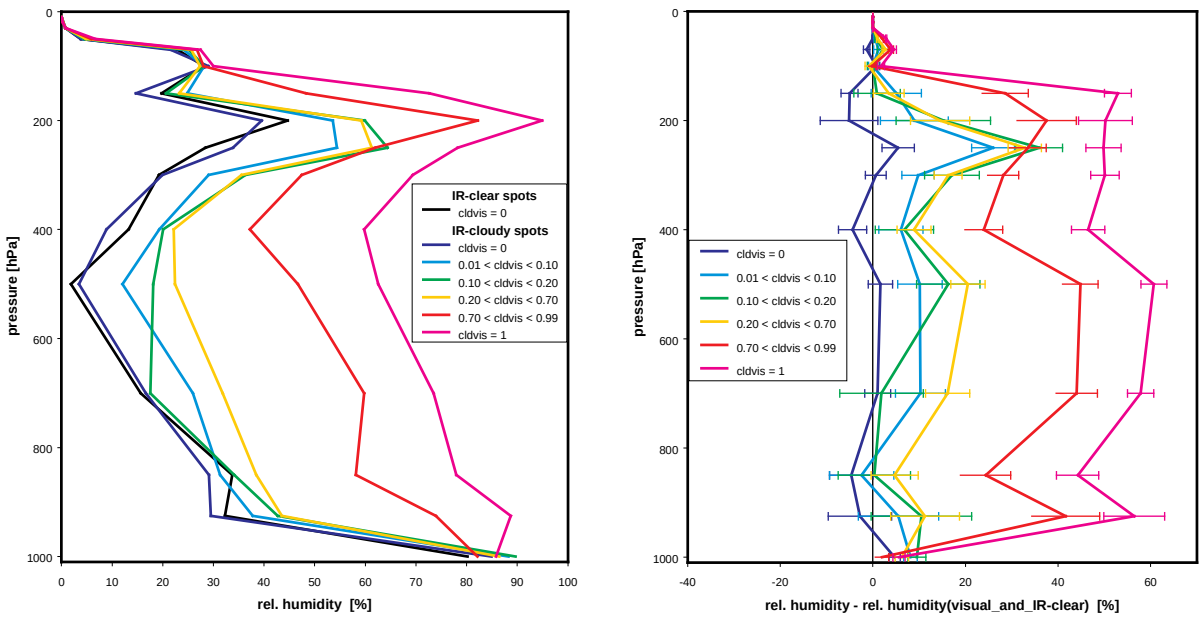

Figure 14: Average model relative humidity profiles for "fully clear" observations (IR-clear and $cld_{vis} = 0$) as well as for several cld_{vis}-classes of IR-cloudy spots (left figure), and the differences of relative humidity with respect to the "fully clear" observations (right figure).

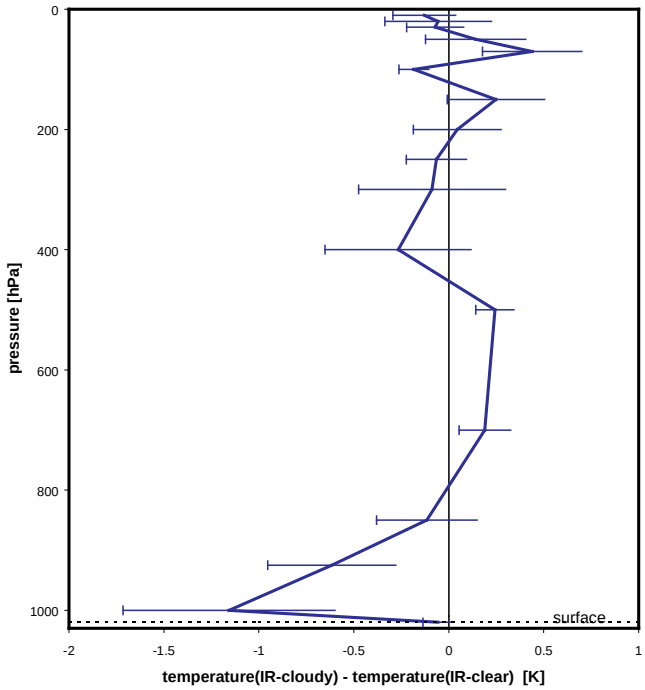

Figure 15: Difference $T_{IR-cloudy}(p) - T_{IR-clear}(p)$ between average model temperature profiles of visual clear ($cld_{vis} = 0$) observations classified as IR-cloudy respectively IR-clear.

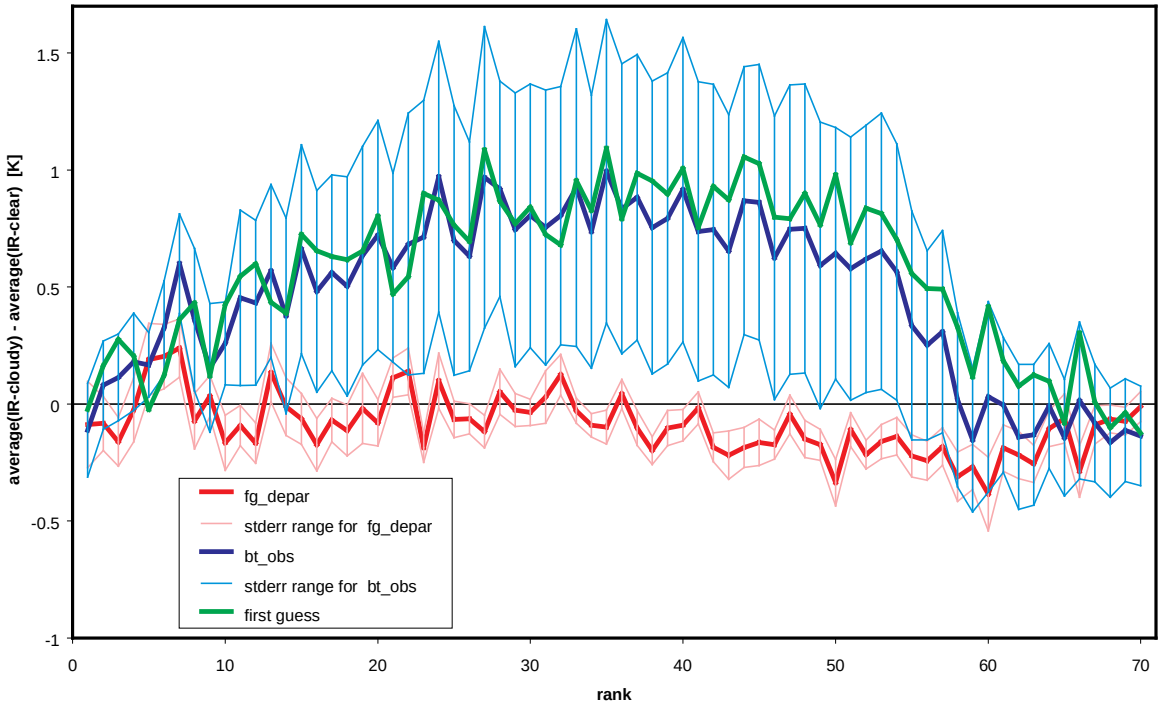

Figure 16: . Differences of average bt_obs, fg_depar, and fg between IR-cloudy and IR-clear visual clear spots for all used tropospheric AIRS channels in the long wave CO_2-band. The channels are rank-ordered according to the height of their sensitivity level (highest rank indicate window channels).

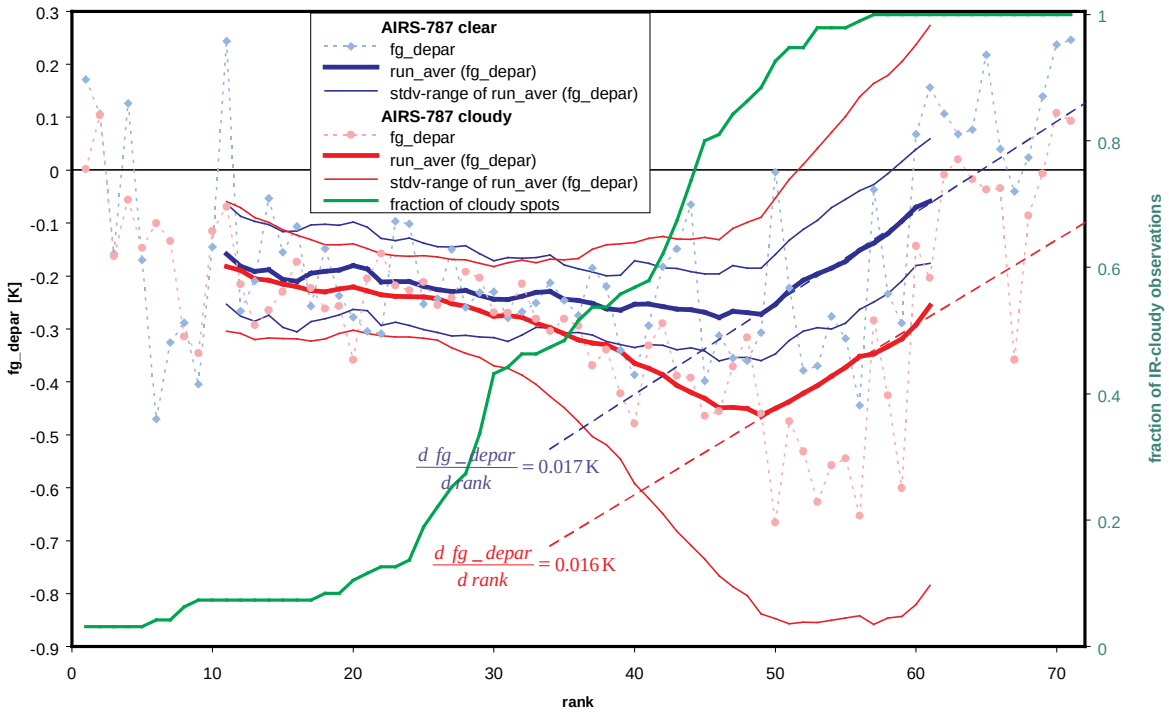

Figure 17: Average fg_depar and its running 21-spots-average in rank-ordered space for two classes of visual clear observations, namely IR-clear and IR-cloudy spots (regarding AIRS-787). The dashed lines show the slope $d\ fg\_depar/d\ rank$ for the near-surface part of the curves.

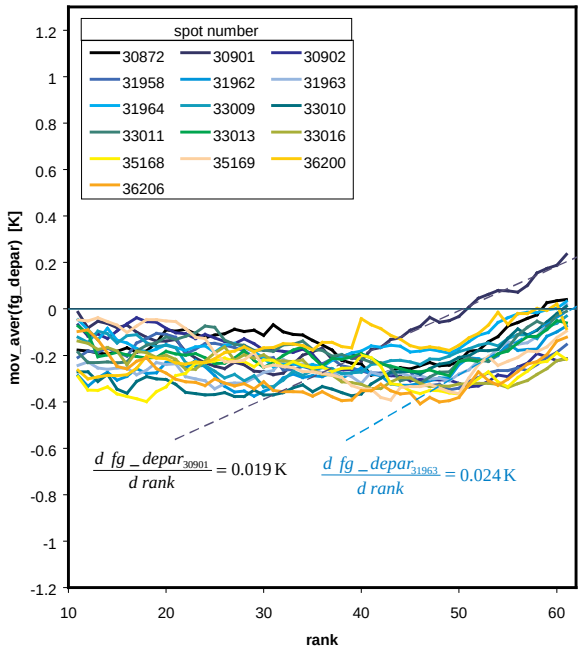

Figure 18: fg_depar curves (running 21-spots-average in rank-ordered space) of all IR-clear observations among the visual clear spots in the investigation area south of Baja California. The dashed lines shows the slopes $d\ fg\_depar/d\ rank$ for the near-surface part of the curves that exhibit the largest impact of cloud or sea surface temperature error respectively showing the steepest slope.

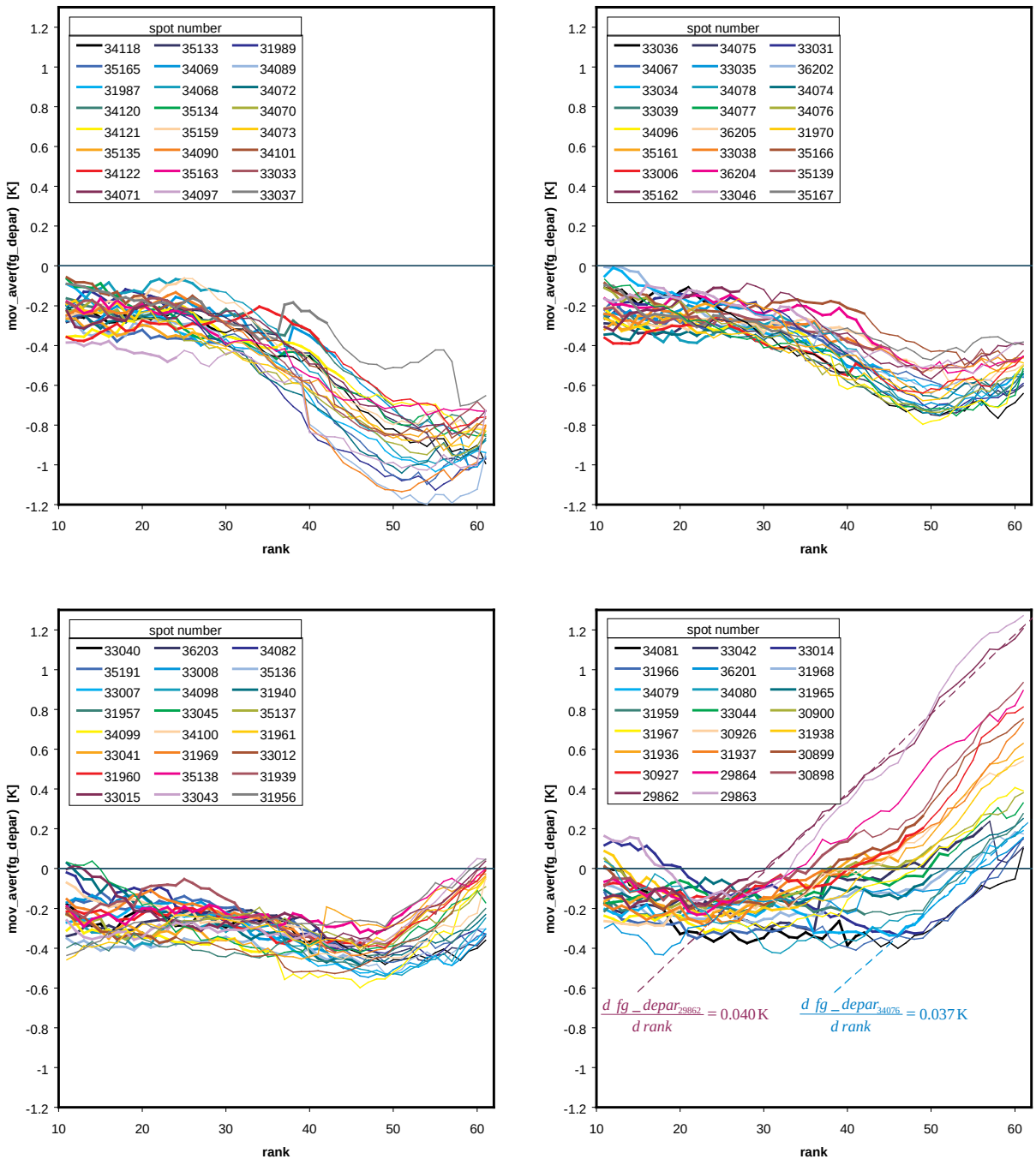

Figure 19: fg_depar curves (running 21-spots-average in rank-ordered space) of all IR-cloudy observations among the visual clear spots in the investigation area south of Baja California. Clear channels are displayed with broad lines, cloudy channels with fine lines. The dashed lines show the slopes $d\ fg\_depar / d\ rank$ for the near-surface part of some typical curves.

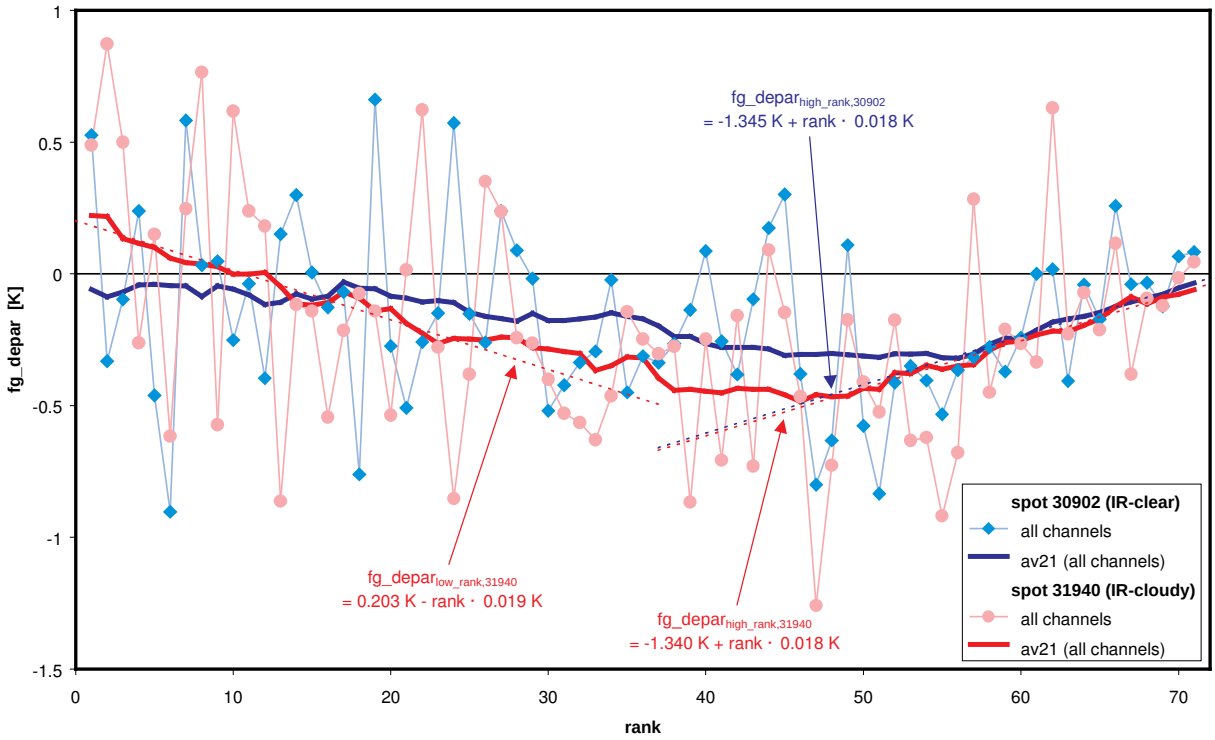

Figure 20: Measured fg_depar for AIRS-787 and its running 21-spots-average in *rank*-ordered space for spot 30902 (IR-clear) and spot 31940 (IR-cloudy). Both spots are visual clear and are located within a distance of less than 20 km of each other. The dotted lines show the slopes $d\,fg\_depar/d\,rank$ for the steepest parts of the average curves.

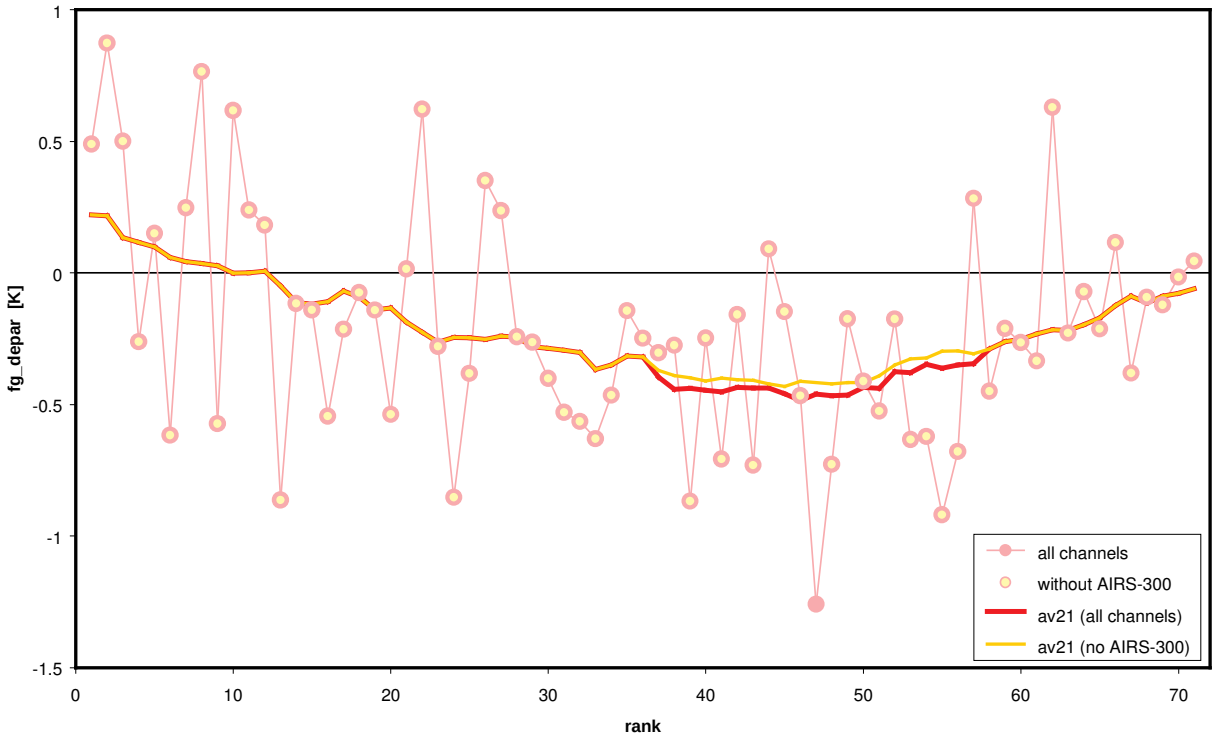

Figure 21: Same as fig.20 for spot 31940 (IR-cloudy) only. The orange curve illustrates the effect of removing the outlier AIRS-300 ($rank = 47$) from the data set. It results in a significant reduction of the depth of the minimum. Note that the original $rank$ is kept in the reduced data set for the sake of better comparability of the curves.

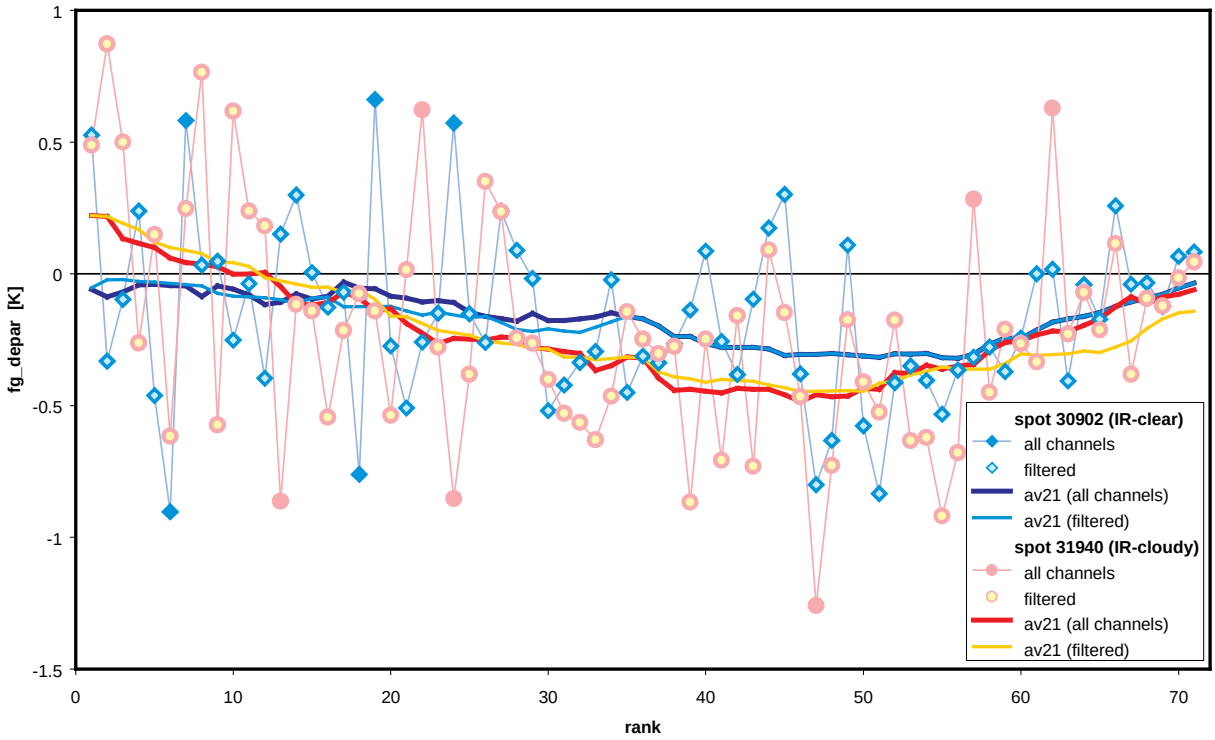

Figure 22: Same as fig.20 illustrating the effect of the outlier elimination with thresholds of $\Delta_{outlier}^{abs} = 0.6\,\mathrm{K}$ and $\Delta_{outlier}^{rel} = 1.5$ for spot 30902 (IR-clear) and spot 31940 (IR-cloudy). As in fig.21 the original *rank* is kept in the reduced data sets for the sake of better comparability of the curves.

5 Summary

The statistical analysis for the AIRS-787 window channel revealed that there are still serious problems with cloud contaminated radiances missed by the cloud detection scheme (type 1 error):

- About 17.7% of the spots passing the cloud detection ("IR-clear spots") are cloud contaminated, resulting in an average "cloud bias" of -0.071 K within all IR-clear observations. The average "cloud bias" of the cloud contaminated IR-clear spots only is approximately -0.4 K.
- The percentage of cloud contaminated spots among all IR-clear observations has a negative correlation to the spatial density of IR-clear spots. **Therefore, the thinning procedure required before data assimilation exhibits a strong preference of cloud contaminated spots. This increases the fraction of contaminated observations within the assimilated data ("used spots") to 40.9% and the resulting "cloud bias" to -0.164 K.**
- As a side effect of these type 1 errors in the cloud detection, a warming excess in the magnitude of 0.04 K in the bias correction for AIRS-787 could be shown, i.e. **clear spots exhibit an average first guess departure of 0.04 K after bias correction**.

- The percentage of cloud contaminated spots among the IR-clear (respectively the used) observations is correlated to the observed visual cloudcover. Thus, the "cloud bias" increases from almost 0 K (resp. -0.86 K) for visual clear spots to -0.142 K (resp. -0.196 K) for spots with a visual cloud cover exceeding 20%.

The case studies focused on both types of possible errors in the cloud detection (missing clouds = type 1 error, rejecting good data = type 2 error) basically yielded the following findings:

- It has been impossible to identify individual cases of type 1 error with a sufficient reliability. Therefore, no other detailed characterisation of features correlated to the occurrence of type 1 errors besides the statistical correlations described above could be found.
- **Many high cirrus clouds clearly visible in the visual spectrum do not affect the AIRS long wave CO_2 band. Thus, these clouds pass the cloud detection. As this is desirable for the purpose of data assimilation, it is not to be judged as type 1 error.**
- The rejection of many visual clear spots by the cloud detection in the 2nd case study turned out to be well-founded. The rejected observations do not only differ significantly from the IR-clear ones with respect to the model conditions, but also with respect to their AIRS radiances (and consequently concerning the true atmospheric state). Although evidence was found for a relevant impact of model errors on the first guess departures in the investigated area, an additional effect of true cloud contamination is also plausible. Thus, type 2 errors could not be proven in this case study.
- Outliers of first guess departure in the ranked channel sequence can affect the cloud detection in a decisive way. This applies particularly to the rejection of observations due to a minimum in the *av21_fg_depar* -curve.

6 Conclusions

As a consequence of the significant percentage of type 1 errors occurring in the cloud detection, a slight cooling of the model atmosphere has to be expected from the assimilation of AIRS radiances, particularly in regions with a relatively low density of IR-clear spots. No indicators of type 1 error in the strict sense could be identified, rather indicators of an above-average probability of type 1 error. Basically, the utilisation of these indicators within a data assimilation system would be possible on two different stages:

- Within the cloud detection: Exclude all spots with a visual cloudcover exceeding a chosen threshold. The most radical solution would be to use visual clear spots only.
- Between cloud detection and thinning procedure: Application of an "inverse thinning", i. e. the elimination of isolated IR-clear spots.

To do so would definitely reduce the fraction of undetected cloud contaminated radiances among the assimilated data. However, this would be obtained at the cost of a huge loss of

good data. Particularly in areas with extended cirrus clouds or with a sparse distribution of clear spots a considerable reduction of usable data has to be faced. Therefore, it is doubtful that a reduction of type 1 errors by these means would be rewarding in the context of a data assimilation system. Nevertheless, experiments should be carried out to figure out this question.

In contrast to the demands of the data assimilation, a huge data loss would do no harm to the calculation of bias correction parameters. Thus, the quality of the bias correction can be improved by applying the methods described above to the data set used for the determination of bias correction parameters.

The lover quality of isolated IR-clear spots (i. e. the high percentage of undetected cloud contamination in these data) raises some questions:

- How can the thinning procedure be optimised to reduce the undesireable preference of cloud contaminated spots?
- Can a different "bias correction for isolated spots" help to reduce the impact of the undetected cloud contamination on the NWP model?
- Are the efforts to increase the number of assimilated AIRS data in cloudy areas by identifying small clear spots ("hole hunting") really promising? How can the risk of poorer data quality associated with hole hunting be reduced?

It has been possible to prove the impact of model errors (particularly temperature errors in the lower troposphere and at the surface) on the first guess departures on a regional scale, i. e. for a sample accessible to statistical methods. Unfortunately, it is much more difficult to distinguish between the impact of model errors or clouds within a single observation. Attributing any occurring distortion in *av21_fg_depar* -curves to an assumed cloud impact will therefore continue to be the only justifiable treatment – although it implies to accept a considerable amount of type 2 errors, especially in areas with serious model errors, i. e. just in those situation, where the assimilation of the AIRS data would be particularly worthwhile.

A reduction of type 2 error frequency as well as a general increase in the quality of assimilated data can be achieved by an improved elimination of outliers. This includes the consequential use of any available data quality information as well as the application of additional outlier-filters based on the *rank* -ordering of the AIRS channels. Further studies are convenient to tune and evaluate the outlier-filter outlined in this report.

The statistical analysis so far has been focused on the AIRS-787 window channel. Further investigations comprising other channels are required to yield a more comprehensive picture of the importance of type 1 errors in the AIRS cloud detection. This applies to the performance in the upper troposphere as well as to the use of humidity information. The latter one requires the analysis of channels from the AIRS water vapour band.

Since the statistical analysis has made use of the visual cloud product, it has been constricted to daytime data. Although one would expect that the cloud detection performance will be similar at nighttime, this issue deserves further investigation. At least the impact of the thinning procedure on the first guess departure statistics should be investigated for nighttime measurements in a similar way as described here.

References

McNally, A. P. and Watts, P. D. (2003): *A cloud detection algorithm for high-spectral-resolution infrared sounders*, Q. J. R. Meteorol. Soc. **129**, 3411-3423

List of used symbols

$av21\_fg\_depar$	running average of first guess departures of 21 subsequent channels in *rank*-ordered space
$av9\_fg\_depar$	running average of first guess departures of 9 subsequent channels in *rank*-ordered space
BC_{exc}	warming excess of bias correction
bt_obs	observed brightness temperature
$bt\_obs_{IR-clear}$	average observed brightness temperature of IR-clear spots
$bt\_obs_{IR-cloudy}$	average observed brightness temperature of IR-cloudy spots
$CB_{dataset}$	"cloud bias" within a data set, i. e. impact of clouds on the average first guess departure of the data set
CB_c	impact of clouds on the average first guess departure of all spots classified as IR-clear
CB_{used}	impact of clouds on the average first guess departure of all spots used for assimilation
CB_{wc}	impact of clouds on the average first guess departure of all spots wrongly classified as IR-clear
cld_{vis}	visual cloudcover (visual cloud product from AQUA satellite)
F_{tc}	fraction of spots correctly classified as IR-clear within a data set
F_{wc}	fraction of spots wrongly classified as IR-clear within a data set: $F_{wc} = 1 - F_{tc}$
fg	first guess, i. e. brightness temperature calculated from model background
$fg_{IR-clear}$	average first guess of IR-clear spots
$fg_{IR-cloudy}$	average first guess of IR-cloudy spots
fg_depar	first guess departure, i. e. observed brightness temperature (bias corrected) minus brightness temperature calculated from model background
$\overline{fg\_depar_c}$	average first guess departure of all spots classified as IR-clear
$\overline{fg\_depar_{dataset}}$	average first guess departure of a data set
$\overline{fg\_depar_{tc}}$	average first guess departure of spots correctly classified as IR-clear

$\overline{fg\_depar_{wc}}$	average first guess departure of spots wrongly classified as IR-clear
p	pressure level
$p_c(fg\_depar)$	frequency distribution of first guess departures for all spots classified as IR-clear
$p_{tc}(fg\_depar)$	frequency distribution of first guess departures for the spots correctly classified as IR-clear
$p_{wc}(fg\_depar)$	frequency distribution of first guess departures for the spots wrongly classified as IR-clear
R_{clear}	calculated outward radiative flux density assuming clear sky conditions
$R_{cloudy}(p)$	calculated outward radiative flux density assuming an opaque black cloud with top at pressure level p
rank	position within a sequence of channels ordered according to their sensitivity level in decreasing order
$RH(p)$	model relative humidity at pressure level p
$RH_{cld_{vis}=0}(p)$	average model relative humidity of visual clear spots at pressure level p
$RH_{IR-clear}(p)$	average model relative humidity of IR-clear spots at pressure level p
$RH_{IR-cloudy}(p)$	average model relative humidity of IR-cloudy spots at pressure level p
stdv	standard deviation
$stdv9\_fg\_depar$	standard deviation of first guess departures of 9 subsequent channels in *rank*-ordered space
$T(p)$	model temperature at pressure level p
$T_{IR-clear}(p)$	average model temperature of IR-clear spots at pressure level p
$T_{IR-cloudy}(p)$	average model temperature of IR-cloudy spots at pressure level p
T_{surf}	model sea surface temperature
$\Delta_{outlier}^{abs}$	threshold for absolute first guess departure deviation defining an "outlier"
$\Delta_{outlier}^{rel}$	threshold for relative first guess departure deviation defining an "outlier"