To: AIRS Design File

From: Thomas Hearty & Evan Manning

CC: E. Fetzer, E. Fishbein, S. Friedman, H. Aumann, B. Lambrigtsen, S-Y. Lee

Subject: Special cases of v4.0.x retrievals: (1) missing forecast surface pressure and (2) regression-only retrieval.

Summary: This memo examines the differences that can be expected when performing two special cases of retrievals with the v4.0.x PGE: (1) retrievals without the surface pressure from the NOAA Global Forecast System (GFS) and (2) regression only retrievals. An understanding of these differences is important for users who may want to give up some accuracy in the retrieval in exchange for a rapid solution (e.g., direct broadcast users).

1 Introduction

This memo examines the characteristics of AIRS v4.0.x Level 2 temperature and water vapor retrievals when the data are processed under either of the 2 following circumstances:

1. Without the forecast pressure from the NOAA Global Forecast System (NOAA/GFS);

2. Only the regression retrieval is performed (i.e., “regression-only”).

The bias and RMS of the AIRS temperature and water vapor retrievals are calculated for both cases with respect to the forecast data from the European Center for Medium-Range Weather Forecasting (ECMWF).

Section 2 examines the differences that are observed when retrievals are processed with and without the NOAA/GFS. Section 3 examines the differences observed when only the regression retrieval is performed. Appendix A lists the Quality Assessment parameters that are not calculated when only the regression retrievals are performed.

Two days are used in this analysis: September 6, 2002 and December 5, 2003. These days are sometimes referred to as “Focus 3” and “Focus 13,” respectively. September 6, 2002 might have have misleadingly better results than December 5, 2003 when compared with ECMWF for two reasons: (1) September 6, 2002 is one of the days used for generating the regression coefficients and (2) The Humidity Sounder of Brazil (HSB) was still operating on September 6, 2002. Therefore, Appendix B redoes the analysis presented in Section 3 and examines September 6, 2002 (Focus 3) data when only the regression retrieval is performed.
2 The Effect of no NOAA/GFS

Each AIRS Infrared channel measures a signal sensitive to temperature, cloudiness, water vapor, and trace gas concentration in a specific range of levels in the atmosphere. Some “window channels” also sense Earth’s surface. Because of the emissive process in the atmosphere, these signals are correlated to a particular pressure altitude (measured in millibars or hectopascals) rather than to a particular elevation in meters above sea level. The precise surface pressure in millibars for a given location varies with passing high- and low-pressure systems which cannot be sensed by AIRS. For those frequencies sensitive to the surface and the lowest part of the atmosphere, the amount of signal coming from the atmosphere versus the surface will shift with changing weather patterns. The different interpretations of the near-surface environment also affect retrieved temperature, water vapor, and trace gas profiles throughout the atmosphere. For this reason, nominal AIRS processing uses a surface pressure derived from a forecast, specifically NOAA’s GFS (formerly called “Aviation Forecast” or “AVN”). When the forecast surface pressure is not available, a climatological average surface pressure is assumed.

It is expected that Direct Broadcast centers may choose to operate without the GFS forecast, both because using the GFS requires pulling significant amounts of data across the Internet in near real time, and because using GFS adds significant complexity to an installation. This section explores the practical differences seen when one day of data (December 5, 2003 – Focus 13) is processed with and without the GFS forecast input.

Figure 1 shows a histogram of the difference between the climatological pressure and the pressure taken from the NOAA/GFS for AIRS “footprints” on December 5, 2003. The pressures for some footprints can differ by more than 40 mbar. The global map shown in Figure 2 demonstrates that the differences in surface pressures are latitude dependent.

Although there are significant differences in the surface pressure, the differences in the retrieved parameters are negligible. The yields (based on the quality assessment parameters $\text{Qual\_Temp\_Profile\_Top} = 0, \text{Qual\_Temp\_Profile\_Mid} = 0, \text{Qual\_Temp\_Profile\_Bot} = 0,$ and $\text{Qual\_H2O} = 0$) for Focus 13 are shown in Figure 3 for AIRS Level 2 data processed with and without the NOAA/GFS. There is almost no difference in the yield between the AIRS retrievals processed with or without the NOAA/GFS. Also, Figure 4 shows that there is no significant difference in the temperature or water vapor retrievals based on the presence of the NOAA/GFS. Figure 5 shows the temperature and water vapor bias and RMS difference between identical footprints from the Focus Day 13 retrievals over ocean that were processed with and without the NOAA/GFS. There is no significant bias between the two cases and the RMS of the difference is on the order of the measurement uncertainty. Similarly, the difference in the retrieved surface temperature (Figure 6) shows a much more normal distribution than that of the surface pressure.
Figure 1: Histograms are displayed for difference between the pressure obtained from the climatology and the NOAA/GFS.

Figure 2: This map shows difference between the surface pressure obtained from the climatology and the NOAA/GFS. The nighttime differences are similar. No day data is present for the North polar region because for this date (December 5, 2003) the region is always in darkness. The vertical crescent-shaped gaps are expected under-lap resulting from the orbital pattern.
Figure 3: The difference in the yield is shown for data processed with and without the NOAA/GFS pressure.
Figure 4: Retrievals with and without the NOAA/GFS are compared to ECMWF data.
Figure 5: Retrievals using a climatological surface pressure ("no GFS") are compared with retrievals that use the NOAA/GFS. There is no significant difference in the two sources of surface pressure.
Figure 6: Histograms are displayed for difference between the retrieved surface temperature obtained using the climatological surface pressure and the NOAA/GFS surface pressure. The temperature difference distribution is much more symmetric than the pressure difference distribution.

3 Regression vs. Final Retrieval

The AIRS v4.0.x retrieval algorithm consists of these major steps:

1. Microwave-Only First Guess
2. Initial Cloud Clearing
3. Initial Regression Retrieval
4. Final Cloud Clearing
5. Final Physical Retrieval

Because most of the processing time is spent in steps 4 & 5, Direct Broadcast stations have the option of stopping after step 3. This section discusses the differences between the products after step 3 (regression-only retrieval) and step 5 (full physical retrieval). The data examined in this section are from December 5, 2003 (Focus Day 13).

On theoretical grounds a full physical retrieval is preferred to a regression retrieval, because regressions can replicate errors and biases in their training sets. The AIRS v4.0.x regression is trained on 3 days of ECMWF data. Since one of these days is September 6, 2002 (Focus Day 3), we examine the results of processing this day in Appendix B.

There are several quantities calculated by the full retrieval that are not present in products of regression-only retrievals. These are listed in Appendix A. Of these, the most serious
loss is of the Qual_* flags, which are usually used to select those retrieved quantities which are of high enough quality for a given purpose. Unfortunately, these parameters are not available with the regression-only retrieval. Users of regression-only retrievals should instead select only those profiles which have retrieval_type = 40 (indicating successful MW-Only retrieval, initial cloud clearing, and regression) and a low value of Initial_CC_score. Initial_CC_Score is an indicator of how well the initial cloud-cleared radiances match radiances reconstructed from clear eigenvectors. It is a unit-less ratio that is described in the Level 2 interface specifications as follows:

1. 0.33 is best possible, a 3× noise reduction
2. < 0.8 is “very good”
3. < 3.0 is “pretty good”
4. > 10 has a “major problem.”

High values typically accompany cases where there were problems with initial cloud clearing or the spectrum may have not well matched any in the training set. Figure 7 shows the distribution of Initial_CC_score. The global map in Figure 8 shows that higher values of Initial_CC_score tend to form coherent regions near places where the regression retrieval failed completely. These are generally regions with the most cloud cover. The yields from regression-only retrievals using various thresholds of Initial_CC_score are displayed in Table 1 and compared with the yields using the standard Quality Assessment parameters calculated by the full retrieval.

The yields for the temperature retrievals for Focus Day 13 are displayed in Figure 9 for the above 3 cases along with the yield based on Qual_Temp_Profile_* = 0 when the full retrieval is performed. The temperature bias and RMS with respect to ECMWF are displayed in Figure 10 for the regression and final retrievals. Figure 11 shows histograms for the difference between the retrieved sea surface temperature (TSurfStd) and the forecast sea surface temperature from ECMWF. The regression-only distributions are more biased

### Table 1: Percent Yield by retrieval_type for December 5, 2003

<table>
<thead>
<tr>
<th>retrieval_type</th>
<th>land</th>
<th></th>
<th>ocean</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number %</td>
<td>Number %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression-only retrieval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>retrieval_type = 40</td>
<td>31018 96.27</td>
<td>123351 96.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial_CC_score &lt; 3</td>
<td>28840 89.51</td>
<td>114946 89.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial_CC_score &lt; 0.8</td>
<td>11434 35.49</td>
<td>58669 45.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>full retrieval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual_Temp_Profile_Top = 0</td>
<td>27307 84.75</td>
<td>112257 87.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual_Temp_Profile_Mid = 0</td>
<td>17991 55.84</td>
<td>74155 57.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual_Temp_Profile_Bot = 0</td>
<td>3889 12.07</td>
<td>58201 45.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual_H2O = 0</td>
<td>26266 81.52</td>
<td>110142 86.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual_Surf = 0</td>
<td>80 0.25</td>
<td>13020 10.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7: The distribution of Initial_CC_score for all of the retrievals with retrieval_type = 40 for a “regression only” run of the v4.0.x PGE on Focus Day 13.

Figure 8: A global map of the daytime retrievals that failed on Focus Day 13, those that have Initial_CC_score > 3, and those that have Initial_CC_score ≤ 3. The distributions for nighttime retrievals and Focus Day 3 retrievals are similar. No day data is present for the North polar region because for this date (December 5) the region is always in darkness. The vertical crescent-shaped gaps are expected under-lap resulting from the orbital pattern.
and not as symmetric as that of the full retrieval but the yield is larger. Also, as the
Initial CC score becomes smaller the bias and the RMS of the difference also decreases.
The water vapor bias and RMS with respect to ECMWF are displayed in Figure 12 for
the regression and final retrievals over ocean and land. The regression-only retrievals with
Initial CC score < 0.8 are, in most cases, as good as the full AIRS retrievals but have a
smaller yield in the upper atmosphere.

4 Recommendations

1. Although retrievals using the NOAA/GFS are preferred, direct broadcast users can
opt to process without NOAA/GFS without significant degradation of their products.

2. Regression retrievals are significantly less reliable than those produced by a full re-
trieval, but are still a viable option when processing speed is paramount. Users must
choose an Initial CC score threshold that balances yield and quality.

A Parameters not calculated in the regression-only re-
trieval

The parameters not calculated by the regression-only retrieval are listed below. A description
of these parameters can be found in the Level 2 interface specifications.

A.1 AIRS Level-2 Standard product

- Qual_*^2
- totO3StdErr
- emisIRStdErr
- rhoIRStdErr
- CC_noise_eff_amp_factor
- CC1_noise_eff_amp_factor
- MW_ret_used

A.2 AIRS Level-2 Support Product

- Qual_*^2
- CO_PPBV, CO_PPBV_Err
- CldEmis, CldEmisErr

^2except Qual_Cloud_OLR & Qual_Guess_Psurf
Figure 9: The yield of the temperature retrievals for Focus Day 13 are displayed for the full v4.0.x retrievals using the `Qual_Temp_Profile_*` parameters and for the “Regression Only” retrieval using various `Initial_CC_score` thresholds.
Figure 10: The bias and RMS of the regression only temperature retrievals are displayed for Focus Days 13.
Figure 11: A histogram showing the difference between the retrieved sea surface temperature and the forecast sea surface temperature from ECMWF for December 5, 2003 (Focus Day 13).

- CldRho, CldRhoErr
- many minor quality indicators (e.g., CC1_Resid, CC1_Noise_Amp, etc.)

A.3 AIRS Level-2 Cloud-Cleared Radiances
- Qual_CC_Rad
- CCR_pass_clear_tests
- CC_noise_eff_amp_factor
- CCfinal_Resid
- clear_lw_resid
- MW_ret_used

B Focus 3 Regression-Only Processing
This section displays figures similar to those shown in Section 3 but for v4.0.x processing of AIRS data from September 6, 2002 (Focus 3). Since this is one of the days used to train the regression coefficients and HSB was still operational at that time, we expect the regression
Figure 12: The bias and RMS of the regression only H$_2$O retrievals are displayed for Focus Days 13.
retrievals will have better agreement with ECMWF data for this day than for the December 5, 2003 (Focus 13) data examined in Section 3.

The yields based on various Quality Assessment parameters are displayed in Table 2. All of the percentages are very similar to that of the values found in Table 1. The yields for the \texttt{Qual.Temp.Profile*} parameters are shown in Figure 13. The bias and RMS difference of the retrieved temperature profiles are displayed with respect to ECMWF data in Figure 14. A histogram showing the difference between the AIRS Sea Surface Temperature and the Sea Surface Temperature from ECMWF is displayed in Figure 15. The bias and RMS with respect to ECMWF is displayed in Figure 16. The yield, bias, and RMS differences of the September 6, 2002 v4.0.x regression retrievals seem to be similar to that of the December 5, 2003 retrievals. However, there are small differences and a more careful analysis should be performed to understand how the data are different.
Figure 13: The yield of the temperature and water vapor retrievals for Focus Day 3 are displayed for the full v4.0.x retrievals using the Qual_Temp_Profile_* parameters and for the “Regression Only” retrieval using various Initial_CC_score thresholds. The yields for land are on top and the yields for ocean are on the bottom.
Figure 14: The bias and RMS of the regression only temperature retrievals are displayed for Focus Days 3.
Figure 15: A histogram showing the difference between the retrieved sea surface temperature and the forecast sea surface temperature from ECMWF for September 6, 2002 (Focus Day 3).
Figure 16: The bias and RMS of the regression only H$_2$O retrievals are displayed for Focus Day 3.