Satellite Application Facility
for Numerical Weather Prediction

Document NWPSAF-MO-UD-022

Version 1.0

January 2010

IASI Principal Components in AAPP: User Manual

Nigel Atkinson

Met Office

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.

Change record			
Version	Date	Author / changed by	Remarks
0.1	10/03/2009	Nigel Atkinson	1st draft
0.2	21/09/2009	Nigel Atkinson	
1.0	20/01/2010	Nigel Atkinson	Initial release

Contents

IASI Principal Components in AAPP: User Manual

1. INTRODUCTION

AAPP version 6.1, released in October 2006, included the ability to convert IASI level 1c radiances to a level 1d product containing Principal Component (PC) scores. A file of reference eigenvectors was supplied, though the user was free to substitute his own eigenvector file if desired. The use of PC scores allows both noise filtering and a large reduction in data volume, compared with the use of raw radiances. See the AAPP Scientific Documentation document for details.

Subsequently the NWP SAF released the "IASI PCA-based compression package" deliverable, which allowed the generation of eigenvectors from a training set, and AAPP was updated to be compatible with the format of the output file from that package (AAPP update 6.6, released Feb 2008).

Since then a new requirement has emerged: the ability to distribute internationally a PC compressed version of the IASI level 1c dataset in the context of EARS-IASI – the IASI component of the EUMETSAT Advanced Retransmission Service. This distribution is required to be in BUFR format. The same format may in the future also be used for IASI near-real-time global data dissemination. The PC BUFR files will contain nominally 290 PCs and 366 channels per spectrum.

Consequently some enhancements to AAPP are required in order to be able to handle this new form of the data. This document describes those enhancements and how they affect the AAPP user.

2. REQUIREMENTS

AAPP is required to fulfill the following requirements related to IASI PCs:

1. Decode the IASI PC BUFR format and store in a format readily ingested by other parts of AAPP (referred to as PC binary format).
2. Create the IASI PC BUFR files starting from PC binary format (i.e. the inverse of requirement 1).
3. Convert the standard IASI level 1c files to IASI PC binary format, using a set of reference eigenvectors. The output is at the full IASI spatial sampling (120 spots per scan).
4. EUMETSAT will distribute reference eigenvector files in HDF5 format, therefore AAPP is required to handle these files. In practice this means converting them to the existing eigenvector files format.
5. Generate reconstructed radiances for any channels that the user selects.
6. ATOVPP currently accepts IASI 1c files as input, together with AMSU and MHS files. It generates an IASI level 1d product (analogous to HIRS 1d). As an alternative to IASI 1c as input, it shall be able to accept IASI PC format files, outputting either reconstructed radiances or the raw radiances (provided the EUMETSAT channel selection contains the required channels).

The flows are illustrated in Figure 1.

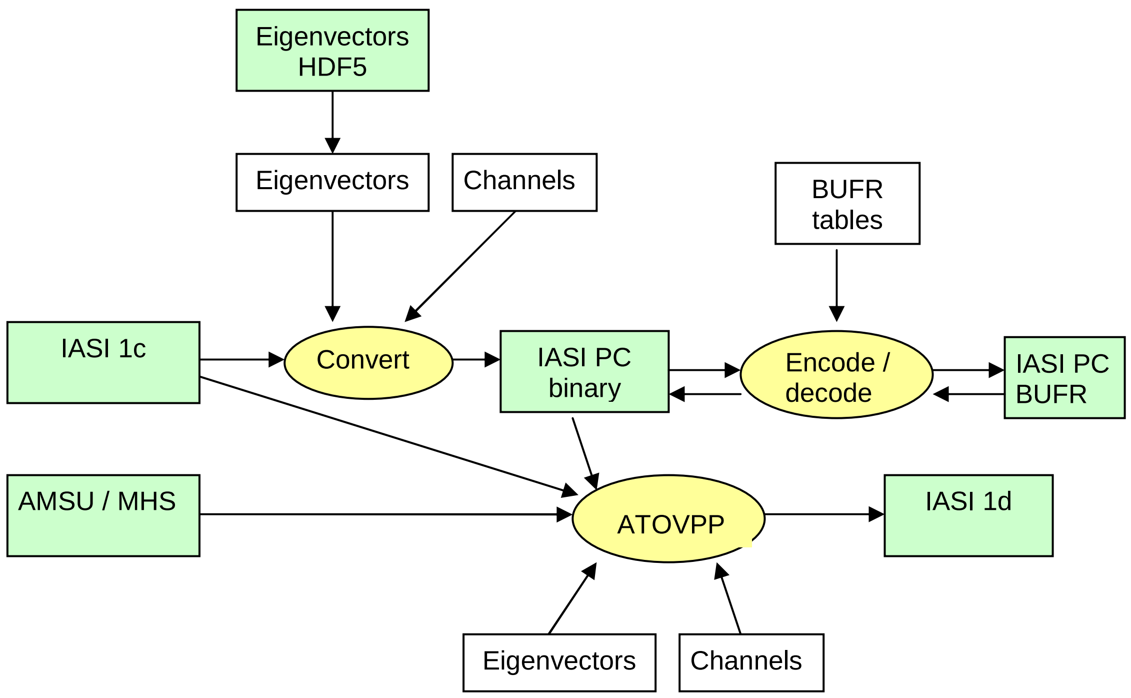

Figure 1: Data flows for IASI PC processing. Processes are in yellow, main data files in green.

3. BUFR ENCODE/DECODE

The executables aapp_encodebufr_1c and aapp_decodebufr_1c have been extended to handle the new formats. Typical commands are:

```
aapp_encodebufr_1c –i iasi.lpc PCIASI
aapp_decodebufr_1c –i iasi.bufr
```

For the encoder, the user has to specify "PCIASI" as the data type; for the decoder, the data type is detected automatically from the sequence code (3-40-008) in the Data Description Section of the BUFR file. The encoder generates an output file with `.bufr` suffix; the decoder generates an output file with `.lpc` suffix.

At the time of writing, the sequence 3-40-008 does not yet appear in the current ECMWF BUFR table D. Similarly there are several additions required to table B (0-40-015 to 0-40-022). Until they are implemented by ECMWF, the BUFR tables will be distributed with AAPP.

As in earlier versions of AAPP, to link the ECMWF BUFR library to the AAPP code, the "configure" command should be used before building AAPP, e.g.

```
BUFR=xxxxx     #point to the directory containing the BUFR library
./configure --external-libs="-L${BUFR}/lib –lbufr" --[other options]
```

As before, at run time the environment variable `$BUFR_TABLES` should point to the directory where the tables are held (with a trailing "/").

4. EIGENVECTOR FILE

For the EARS-IASI service, reference eigenvector files will be distributed by EUMETSAT. A separate file will be used for each of the three IASI spectral bands. The file names will follow the standard EPS file naming convention, and the initial set will be distributed with AAPP:

```
IASI_EV1_xx_M02_20000101000000Z_xxxxxxxxxxxxxxZ_20100113000102Z_xxxx_xxxxxxxx90
IASI_EV2_xx_M02_20000101000000Z_xxxxxxxxxxxxxxZ_20100113000102Z_xxxx_xxxxxxx120
IASI_EV3_xx_M02_20000101000000Z_xxxxxxxxxxxxxxZ_20100113000102Z_xxxx_xxxxxxxx80
```

Each file contains three dates: start time for validity, end time for validity (not specified in the example above) and product generation time. The "hour/minute" fields of the product generation time are by EUMETSAT convention used to store a version number, which should correspond to the "database identification" number in the BUFR sequence. Thus the user can check that a received BUFR file is consistent with the expected eigenvector files. In the example above, 000102 means version 1.02. The final field indicates the number of eigenvectors supplied for each band.

To work with these files the user will need to install the HDF5 library and link to it while building AAPP. The current release (1.8.2 at the time of writing) may be downloaded from http://www.hdfgroup.org/HDF5/index.html . Either the source code or pre-built binaries may be used. (It is easier to use the binaries). You will also need the szip compression library, as explained on the web page.

If your system is not already set up to use HDF5 then you may need to set up your `$LD_LIBRARY_PATH` environment variable at run time, and when building AAPP, as in the following example:

```
HDF5=~/hdf5/5-1.8.0-linux
SZIP=~/hdf5/szip-2.1/szip
for lib in $HDF5 $SZIP; do
  if [[ "$LD_LIBRARY_PATH" != *${lib}* ]]; then
    LD_LIBRARY_PATH=$LD_LIBRARY_PATH:${lib}/lib
  fi
done
```

This code fragment has been added to the default ATOVS_ENV6 file, and may be uncommented and customized as required.

HDF5 utilities such as h5dump can be found in `${HDF5}/bin`. You may wish to add this directory to `$PATH`.

Before building AAPP you should run the "configure" command as follows:

```
./configure \
  --external-libs="-L${BUFR}/lib -lbufr -L${HDF5}/lib -lhdf5" \
  --external-includes="-I${HDF5}/include" \
  --[other options]
```

It is assumed in the above that you also require the BUFR library. Note the new

`--external-includes` option in the configure script.

To set up AAPP to use a particular set of files, edit the file `iasi_eigenvectors_spec.dat` in directory `${DIR_PREPROC}`. You will need to enter the names of the three eigenvector files and the number of PCs to use for each (recommended 90, 120, 80). You can, if you wish, change the name and/or location of this file by setting environment variables `$IASI_FDF` and `$DIR_IASI_PREPROC`.

To display an ASCII dump of the eigenvectors you can run the AAPP command

```
iasi_ev_readhdf5.exe filename number_of_PCs
```

The output is in the same ASCII format as the ECMWF PCA package.

The binary files required by AAPP are created automatically by an executable *iasi_eigenvectors*. You do not need to run this manually as it is called by the other scripts when needed.

To re-generate the binary files, just re-edit (or "touch") the `iasi_eigenvectors_spec.dat` file. They will re-generated next time they are needed.

Note that eigenvector files supplied by EUMETSAT will be based on Gaussian apodised (1C) spectra, not self-apodised (1B). Theoretically one would expect that the use of self-apodised eigenvectors would allow a smaller number of PCs to be used, because self-apodised spectra have diagonal noise covariance. However in practice the reduction has been found to be small, and does not justify the increased computation complexity of additional apodisation-deapodisation cycles. AAPP can handle self-apodised eigenvectors if necessary.

5. SPECIFYING THE CHANNEL SELECTION

The channel selection is normally held in the file `${DIR_IASI_PREPROC}/IASI.fdf`. The ECMWF 366 channel set is recommended to be used for the PC product. The list of channels is available in the "EARS Operational Service Specification". There are 211 channels in band 1, 106 in band 2 and 49 in band 3.

If you wish to specify a different channel selection file, then this can be done using the environment variable `IASI_FDF` (see below). The old 314-channel set may be specified as an alternative.

By default, AAPP allocates storage for up to 366 channels in the PC file. If more are required then you will need to edit file *iasipc.h* and re-build AAPP.

6. CONVERTING FROM 1C TO PC FORMAT

To convert from AAPP 1c binary format to PC binary format, run the following AAPP command:

```
iasi_1c_to_pc infile outfile
```

It is recommended to give the output file a `.lpc` suffix. The command uses the following environment variables:

`$IASI_SPECFILE` File containing eigenvectors spec
(defaults to `iasi_eigenvectors_spec.dat`)
`$IASI_FDF` File containing list of channels to reconstruct,
(defaults to `IASI.fdf`)
`$DIR_IASI_PREPROC` defaults to `$DIR_PREPROC`

7. GENERATING RECONSTRUCTED RADIANCES IN THE PC FILE

The IASI PC binary file includes storage for reconstructed radiances. To generate these, run the following AAPP command:

```
iasi_reconstruct pcfile
```

The environment variables used are the same as those listed in Section 6. Obviously the eigenvector files used in the reconstruction must be the same as the ones used when the PC scores were originally generated: the program checks that the version number recorded in *pcfile* is consistent with the eigenvector file names (from `iasi_eigenvectors_spec.dat`), and aborts if they are not consistent.

By default, AAPP allocates storage for up to 366 reconstructed channels in the PC file. If more are required then you will need to edit file *iasipc.h* and re-build AAPP.

8. USING THE PC FILE IN ATOVPP

To use the IASI PC file as an input to ATOVPP, first re-name (or link) the input file to `iasi.lpc`. Then specify `PCIASI` in the list of instruments, e.g.

```
atovpp -i ”PCIASI AMSU MHS” –g IASI
```

The following environment variables are used by ATOVPP:

`$IASI_SPECFILE` File containing eigenvectors spec
(defaults to `iasi_eigenvectors_spec.dat`)
`$IASI_FDF` defaults to `IASI.fdf`
`$DIR_IASI_PREPROC` defaults to `$DIR_PREPROC`
`$ATOVPP_USE_RR` Set =T (or =TRUE) if reconstructed radiances are to be used.

The channel selection and spatial thinning method are specified in
`${DIR_IASI_PREPROC}/IASI.fdf`.

The channels and PC scores in the output file are selected according to Table 1 below. To output reconstructed radiances, first run *iasi_reconstruct* as described in Section 7.

Table 1: Channels and PC scores to be output by ATOVPP

Input data	Level 1c (full spectrum)	PC format (PC scores + channels)
Output PC scores	Generated by ATOVPP using the supplied eigenvectors	Copied from the input file
Output channels	Raw scaled radiances, selected according to IASI.fdf	Either raw radiances (if requested channels are present in PC file) or reconstructed radiances, depending on $ATOVPP_USE_RR

9. EXAMPLES

Example 1: EARS-IASI ground stations

In this example the *AAPP_RUN_METOP* script is used to process MetOp Level 0 data in directory `$IN`, with output files placed in directory `$OUT`. Processing takes place in `${WRK}`. For each instrument and processing level there is one output file per satellite overpass. Then create BUFR output products, also in `$OUT`.

It is assumed that HDF5 eigenvector files are provided as input, so the *iasi_eigenvectors_spec.dat* file must first be prepared as indicated in Section 4. The 366 channels are held in `IASI.fdf`.

```
# Start with an empty output directory
cd $WRK
rm -f ${OUT}/*

# Process the ATOVS data to level1c, and generate a hirs 1d file
# with AVHRR cloud mask
AAPP_RUN_METOP -i "AMSU-A MHS HIRS AVHRR" -g "HIRS" -d $IN -o $OUT

# BUFR encode the ATOVS data
if [ $? = 0 ]; then
  afile=${OUT}/amsua*.l1c
  aapp_encodebufr_1c -i $afile AMSU-A
  mfile=${OUT}/mhs*.l1c
  aapp_encodebufr_1c -i $mfile MHS
  hfile=${OUT}/hirs*.l1c
  aapp_encodebufr_1c -i $hfile HIRS
fi

# Process IASI to level 1c (using OPS-LRS)
AAPP_RUN_METOP -i IASI -g " " -d $IN -o $OUT

# Generate IASI PC scores and BUFR encode
if [ $? = 0 ]; then
  ifile=${OUT}/iasi*.l1c
  pcfile=$(echo $ifile | sed s/iasil1c/iasilpc/ | sed s/.l1c/.lpc/)
  iasi_1c_to_pc $ifile $pcfile
  aapp_encodebufr_1c -i $pcfile PCIASI
fi

# Remove processed files from the ${IN} directory for next pass
```

```
# (or move to archive)
rm -f ${IN}/*
```

A typical output file name is

`iasilpc_M02_20090311_0845_12412.bufr`

In practice EUMETSAT would re-name this using the standard EARS file naming convention before dissemination to users.

Example 2: ATOVPP users

This might include users of the IASI/AMSU/MHS global data who wish to make use of EARS-IASI.

The incoming BUFR files for AMSU, MHS and IASI are in directory `$IN`. The user generates reconstructed radiances for his own channel selection (which may be different from the standard 366 channel set) and creates an IASI level 1d file in directory `$OUT`. The working directory is `$WRK`.

```
# Remove unwanted files
cd $WRK
rm -f *.l?c ${IN}/*.l?c

# Decode the BUFR files
aapp_decodebufr_1c -i "${IN}/*.bufr"
mv ${IN}/*.l?c $WRK

# Link to files
ln -sf amsua*.l1c aman.l1c
ln -sf mhs*.l1c ambn.l1c
ln -sf iasi*.lpc iasi.lpc

# Compute reconstructed radiances
export IASI_FDF=IASI.fdf   #file containing channel selection
iasi_reconstruct iasi.lpc

# Run ATOVPP
export ATOVPP_USE_RR=TRUE
atovpp -i "AMSU-A MHS PCIASI" -g "IASI"

if [ $? = 0 ]; then
  ifile=$(ls iasi?*.lpc)
  outfile=${ifile%.lpc}.l1d
  mv iasi.l1d ${OUT}/${outfile}
  echo "created $outfile"
else
  echo "Error in atovpp"
fi
```

10. THE USE OF OPS-LRS "DAY 2" AND ENHANCED LEVEL 1C FILES

The Principal Components modules prepared for EARS-IASI will be released to AAPP users as part of AAPP update 6.12.

Independently of the EARS-IASI developments, a new version of OPS (referred to as "day 2") has been prepared by CNES for use in the EPS Core Ground Segment. This is being made available to AAPP users as OPS-LRS version 5-0. It is proposed to release OPS-LRS version 5-0 and AAPP update 6.12 to users at the same time.

Note that whilst AAPP 6.12 can be run with the old version of OPS-LRS (version 4-0), the new version of OPS-LRS requires the AAPP library updates that are in v6.12. So it is recommended that users first install v6.12 and then the new OPS-LRS.

The new OPS-LRS uses an updated level 1c PFS format, which contains certain data not previously available:

- Additional quality flags, including a flag that indicates which of the three IASI bands have good data (previously a single flag covered all three bands).
- Information on the mean and variance of the IASI Imager radiance, for each IASI spot.
- Cloud amount, based on AVHRR
- A snow/ice fraction, based on AVHRR
- Land fraction

The cloud fraction, snow fraction and land fraction are not generated by OPS-LRS, the values come from elsewhere in the Core Ground Segment, and whilst the values will be available to users of the EUMETSAT global data they will not initially be available to direct readout users.

The new OPS-LRS also includes a default value for the IASI interferometer axis when used with direct readout data, which should give improved local-global consistency.

To convey this new information to users of the IASI global 1c data (all channels), a revised BUFR sequence will be used by EUMETSAT, 3-40-007 (see Appendix). Also the internal AAPP IASI level 1c format has been modified (see `include/iasi.l1c` in the AAPP distribution).

AAPP users can, if they wish, create BUFR files using the 3-40-007 sequence, by setting the environment variable `ENHANCED_IASI=Y` . Then use the standard *aapp_encodebufr_1c* tool with the instrument specified as IASI. If `ENHANCED_IASI` is undefined or set to `N` then the old sequence 3-40-001 will be used.

For transmission of a reduced channel set on the GTS, EUMETSAT will be using the 3-40-008 sequence. This is the same sequence as will be used for EARS-IASI, but the number of channels and PC scores will be different (at least initially):

- GTS data will have 300 channels and no PC scores
- EARS-IASI data will have 366 channels and 290 PC scores

The BUFR sequences are shown in the Appendix.

APPENDIX: BUFR SEQUENCES FOR EARS-IASI AND "ENHANCED" GLOBAL DATA

Sequence descriptor (3-40-008): IASI Level 1c data (subset of channels and PC scores)

Descriptor	Data index	Description
		Satellite processing information
0-01-007	1	Satellite identifier
0-01-031	2	Identification of originating/generating centre
0-02-019	3	Satellite instruments
0-02-020	4	Satellite classification
		Date and time
0-04-001	5	Year
0-04-002	6	Month
0-04-003	7	Day
0-04-004	8	Hour
0-04-005	9	Minute
2-02-131		Add 3 to scale
2-01-138		Add 10 to width
0-04-006	10	Second
2-01-000		Reset width
2-02-000		Reset scale
		Location information
0-05-001	11	Latitude (high accuracy)
0-06-001	12	Longitude (high accuracy)
0-07-024	13	Satellite zenith angle
0-05-021	14	Bearing or azimuth
0-07-025	15	Solar zenith angle
0-05-022	16	Solar azimuth
0-05-043	17	Field of view number
0-05-040	18	Orbit number
2-01-133		Add 5 to width
0-05-041	19	Scan line number
2-01-000		Reset width
2-01-132		Add 4 to width
0-25-070	20	Major frame count
2-01-000		Reset width
2-02-126		Subtract 2 from scale
0-07-001	21	Height of station
2-02-000		Reset scale
		Quality information
1-03-003		Repeat next 3 descriptor 3 times
0-25-140	22,25,28	Start channel
0-25-141	23,26,29	End channel
0-33-060	24,27,30	GQisFlagQual
0-33-061	31	QGisQualIndex
0-33-062	32	QGisQualIndexLoc
0-33-063	33	QGisQualIndexRad
0-33-064	34	QGisQualIndexSpect
0-33-065	35	GQisSysTecSondQual
0-40-020	36	*GQisFlagQualDetailed - Quality flag for the system*
		IASI subset of channels

1-01-010		Repeat next 1 descriptor 10 times
3-40-002	37-66	IASI Level 1c band description
1-04-000		Delayed replication of next 4 descriptors
0-31-002	67	Extended delayed replication factor
2-01-136		Add 8 to width
0-05-042	68,70,…	Channel number
2-01-000		Reset width
0-14-046	69,71,…	Scaled IASI radiance
		Instrument band definition
1-08-003		Repeat next 8 descriptors 3 times
0-25-140	68+2n,…	Start channel
0-25-141	69+2n,…	End channel
0-40-015	*70+2n,…*	*Quantization factor*
0-40-016	*71+2n,…*	*Residual RMS in band*
0-25-062	72+2n,…	Database identification
		Principal component scores for band
1-01-000		Delayed replication of one descriptor
0-31-002	73+2n,…	Extended delayed replication factor
0-40-017	74+2n,…	*Non-normalised principal component score*
		AVHRR scene analysis
0-02-019	86+2n+p	Satellite instruments
0-25-051	87+2n+p	AVHRR channel combination
1-01-007		Repeat next 1 descriptor 7 times
3-40-004	88+2n+p,…	IASI Level 1c AVHRR single scene sequence
0-20-081	319+2n+p	Cloud amount in segment
0-08-029	320+2n+p	Remotely sensed surface type
0-20-083	321+2n+p	Amount of segment covered by scene
0-08-029	322+2n+p	Remotely sensed surface type
0-40-018	323+2n+p	*Average of imager measurements*
0-40-019	324+2n+p	*Variance of imager measurements*
0-40-021	325+2n+p	*Fraction of weighted AVHRR pixel in IASI FOV covered with snow/ice*
0-40-022	326+2n+p	*Number of missing, bad or failed AVHRR pixels*

where n is the number of channels and p is the number of PCs.
New descriptors are in italics.
Descriptors related to “enhanced” IASI information are underlined. They will be set to “missing” for OPS-LRS version 4-2 (operational in 2009) and all earlier versions.

Sequence descriptor (3-40-007): IASI Level 1c data (all channels)

Descriptor	Data index	Description
		Same as 3-40-008 up to here …
0-40-020	36	*GQisFlagQualDetailed - Quality flag for the system*
		IASI channels
1-01-010		Repeat next 1 descriptor 10 times
3-40-002	37	IASI Level 1c band description
1-01-087		Repeat next 1 descriptor 87 times
3-40-003	67	IASI Level1c 100 channel sequence
		AVHRR scene analysis
0-02-019	17467	Satellite instruments
0-25-051	17468	AVHRR channel combination
1-01-007		Repeat next 1 descriptor 7 times
3-40-004	17469	IASI Level 1c AVHRR single scene sequence
0-20-081	17700	Cloud amount in segment
0-08-029	17701	Remotely sensed surface type
0-20-083	17702	Amount of segment covered by scene
0-08-029	17703	Remotely sensed surface type
0-40-018	17704	*Average of imager measurements*
0-40-019	17705	*Variance of imager measurements*
0-40-021	17706	*Fraction of weighted AVHRR pixel in IASI FOV covered with snow/ice*
0-40-022	17707	*Number of missing, bad or failed AVHRR pixels*

New element descriptors:

Descriptor	Name	Units	Scale	Reference	Width
0-40-015	Quantization factor	Numeric	2	0	16
0-40-016	Residual RMS in band	Numeric	3	0	14
0-40-017	Non-normalised Principal Component score	Numeric	0	-1073741824	31
0-40-018	GIacAvgImagIIS - Average of imager measurements	$W/(m^2\ sr\ m^{-1})$	6	0	24
0-40-019	GIacVarImagIIS - Variance of imager measurements	$W/(m^2\ sr\ m^{-1})$	6	0	24
0-40-020	GQisFlagQualDetailed - Quality flag for the system	Flag Table	0	0	17
0-40-021	Fraction of weighted AVHRR pixel in IASI FOV covered with snow/ice	%	0	0	7
0-40-022	Number of missing, bad or failed AVHRR pixels	Numeric	0	0	7

New Flag Table (0-40-020) " GQisFlagQualDetailed - Quality flag for the system "

Bit number	Meaning
1	NZPD and Complex calibration error
2	Band 3 affected by spike
3	Band 3 affected by saturation
4	Band 2 affected by spike
5	Band 1 affected by spike
6	Overflow/Under Flow
7	On-board processing error
8	Spectral calibration error
9	Radiometric calibration error
10	Missing AVHRR data
11	Missing IIS data
12	Missing Sounder data
13	GQisFlagQual summary flag for all bands
14	On-Ground processing error
15	Inter-calibration error IASI/AVHRR
16	Spare
All 17	Missing