

ERA-40 Project Report Series

21. The use of VTPR raw radiances in ERA-40

X. Li, G. Kelly, S. Uppala, R. Saunders and J.K. Gibson

Series: ECMWF ERA-40 Project Report Series

A full list of ECMWF Publications can be found on our web site under:
<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

© Copyright 2005

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

The use of VTPR raw radiances in ERA-40

X. Li¹, G. Kelly, S. Uppala, R. Saunders²
and J.K. Gibson

December 2005

Current affiliation:

¹ WWB, EMC/NCEP/NOAA, Camp Spring, MD, USA

² Met Office, Exeter

1. Introduction

Measurements from the Vertical Temperature Profile Radiometer (VTPR) launched on the NOAA polar orbiters (NOAA-2 to 5) are available from November 1972 to February 1979, and represent a valuable source of the global top of atmosphere radiance observations for reanalyses. Previously retrieved VTPR temperature and humidity profile products have been used in the NCEP/ NCAR reanalysis from 1975 onwards (Kalnay et al. 1996). The horizontal spacing was poor (about 600km). Their quality was not good enough to use before 1975.

In the ERA-40 reanalysis, Uppala et al. (2005), the VTPR raw radiance data have been extensively used. They offered the first long term record of top of atmosphere infrared radiances and they were assimilated from 1973 onwards. Hence the year 1973 represents the beginning of the use of satellite data in ERA-40. The next significant change to the observing system was introduced by the TOVS (TIROS Operational Vertical Sounder) radiance data in 1979, Hernandez et al. (2004).

2. Basic characteristics of the VTPR

VTPR was an operational 8-channel sounding instrument onboard the satellites from NOAA-2 to 5, each spacecraft having two VTPR instruments on board. The satellites with VTPR instruments are shown in Figure 1.

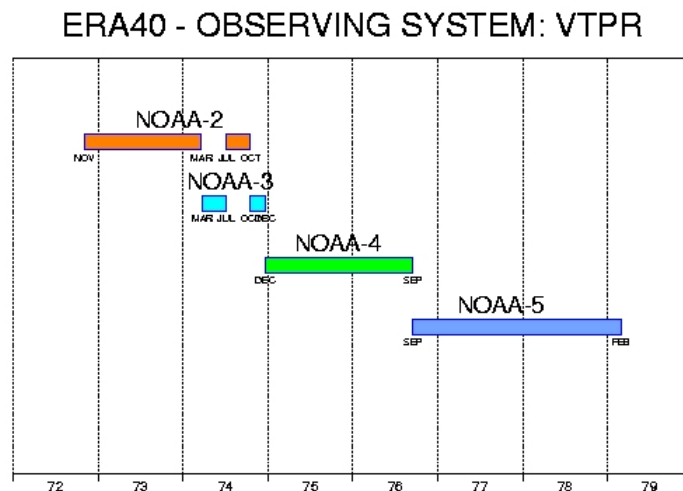


Figure 1. Availability of the VTPR data.

The characteristics of VTPR instrument are summarized in Tables 1 and 2. Six of the eight channels are in the 15 micron CO₂ band. They are very similar to the HIRS CO₂ channels. Two other channels are in the 18 micron water vapour absorption band and at 11 micron in an atmospheric window region. These two channels are used to determine the upper atmospheric water vapour amount and the surface or the cloud top temperature. The weighting functions, showing regions of the atmospheric sensitivity are shown in Figure 2.

Table 1. The VTPR System Parameters

Parameter	Value
Ground field of view (FOV)	
Nadir	55 km along track × 57 km cross track
End of scan	67 km along track × 91 km cross track
Scan line	
Time per scan	12.5 s
Angle from nadir	± 30.3°
Number of FOV	23
Scan step	
Time	0.5 s
Angle	2.7°
Number of channels	8

Table 2. Nominal Spectral Intervals for the VTPR Filters

Channel	Description	Central Wave number (1/cm)	Central Wave length (μm)	Half-power band width (1/cm)
1	15 μm CO ₂	669	14.96	3.5
2	15 μm CO ₂	678	14.77	10
3	15 μm CO ₂	695	14.38	10
4	15 μm CO ₂	708	14.12	10
5	15 μm CO ₂	725	13.79	10
6	15 μm CO ₂	747	13.38	10
7	18 μm H ₂ O	535	18.69	10
8	11 μm Window	833	11.97	10

VTPR Radiance Sensitivity

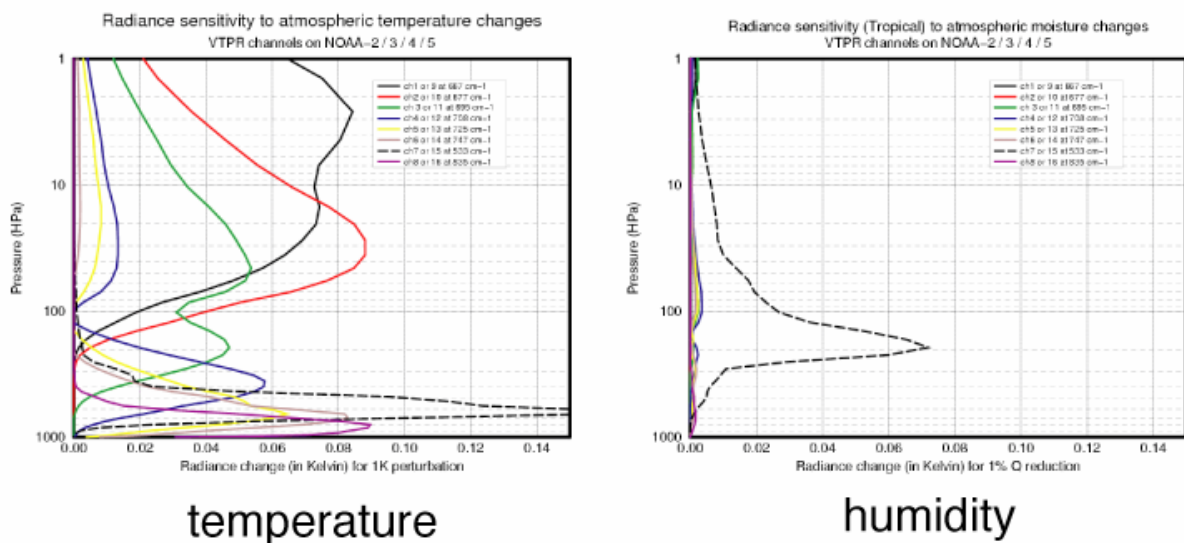


Figure 2 Atmospheric Sensitivity for VTPR channels

3. Pre-processing the VTPR data set

Copies of the level-1b VTPR raw data were obtained from NCAR. These data were recorded many years ago and had not been processed to level-1c. There were many anomalous values of radiance counts in the data stream and considerable effort was required to extract good calibrated brightness temperatures. Details are given in the Appendix A.

Only a few of the VTPR pixels were earth-located. In an area with 184 (8 x 23) pixels only three of them, the "principal points" and coloured red in Figure 3, are available in the original VTPR dataset. The first operational satellite retrievals were produced by a statistical method and used in the NCEP/ NCAR reanalysis from 1975 onwards.

Navigation was required for all the VTPR pixels. By using the ITPP5.01 software all the geographical coordinates were calculated and checked for consistency with the original VTPR data at the same pixel locations, see Figure 4. In addition the epoch times were often incorrect in the original VTPR dataset. The files of orbit parameters, orbital elements, were recreated for NOAA-2 → NOAA-5.

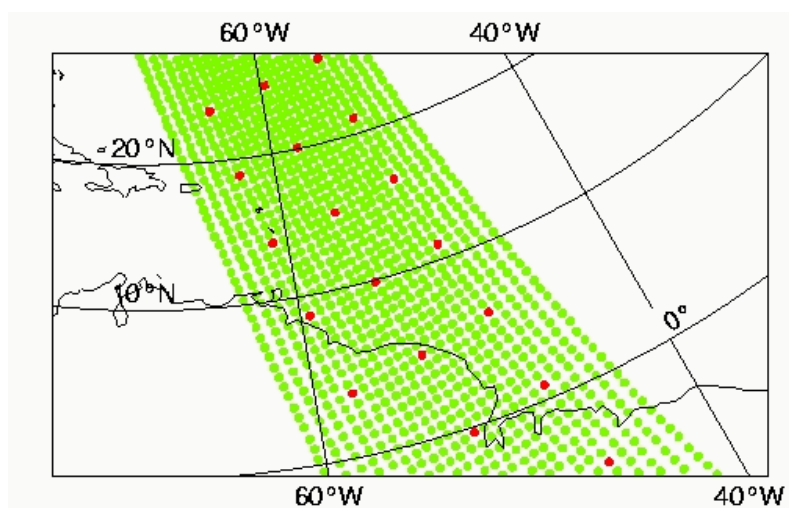


Figure 3. Diagram of the VTPR scan pattern and the "principal points". Only the principal points, the pixels in red, are earth-located in the original VTPR dataset.

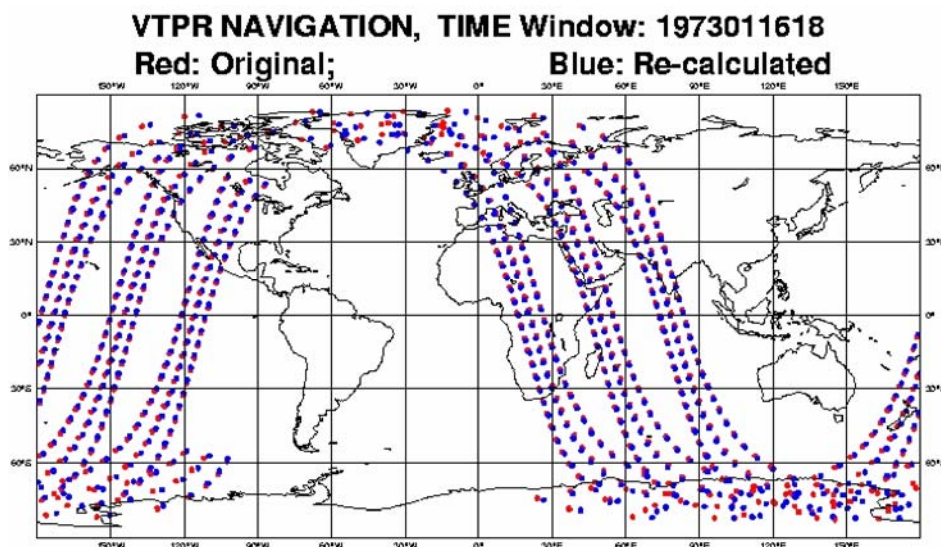


Figure 4. Results of recalculating the navigation for VTPR.

4. Radiance simulator for VTPR

In order to assimilate the VTPR radiances in the ERA-40 3D-Var analysis they have to be simulated using a fast radiative transfer model. In common with other passive infrared and microwave sensors used in ERA-40 the VTPR radiances are simulated using the fast radiative transfer model, RTTOV, described by Saunders et al (1999). In order to compute the VTPR transmittances required for the RTTOV coefficient generation the VTPR channel spectral responses are required. The spectral responses for VTPR are available in McMillin et al (1973) but it proved impossible to relate the different responses to a specific NOAA platform and so the responses were assumed to be the same for all satellites which is clearly not ideal. As there are two sensors on each NOAA platform and the sensor being used is identified in the data stream the coefficients for VTPR1 or VTPR2 could be selected for the ERA-40 radiance simulations. Having computed the transmittances convolved with the filter responses the RTTOV coefficients were computed allowing rapid calculations to be made of VTPR radiances given an input state vector from the model background field. The VTPR RTTOV coefficients are available on request from the NWP Satellite Application Facility at nwpsaf@metoffice.gov.uk for all NOAA platforms (although they are the same coefficients) and for sensor 1 and sensor 2 on each platform.

5. Cloud screening

The method to identify VTPR radiances unaffected by cloud was based on the use of threshold values which varied with channel and latitude. A table of departure values was obtained by comparing the 11 micron window channel (channel-8) brightness temperatures with those calculated from the model first guess by using the radiative transfer model. Figure 5 shows the channel usage above various cloud tops, the cloud height being determined from the channel-8 departure in the table. Additionally a first guess check (observation – background) and variational quality control were applied during the assimilation process to remove bad or inconsistent data. Figure 6 shows a plot of raw radiances for channel-2 without any corrections. This channel peaks in the stratosphere and is unaffected by cloud. Figure 7 shows a typical six hour coverage of VTPR radiances used in the analysis.

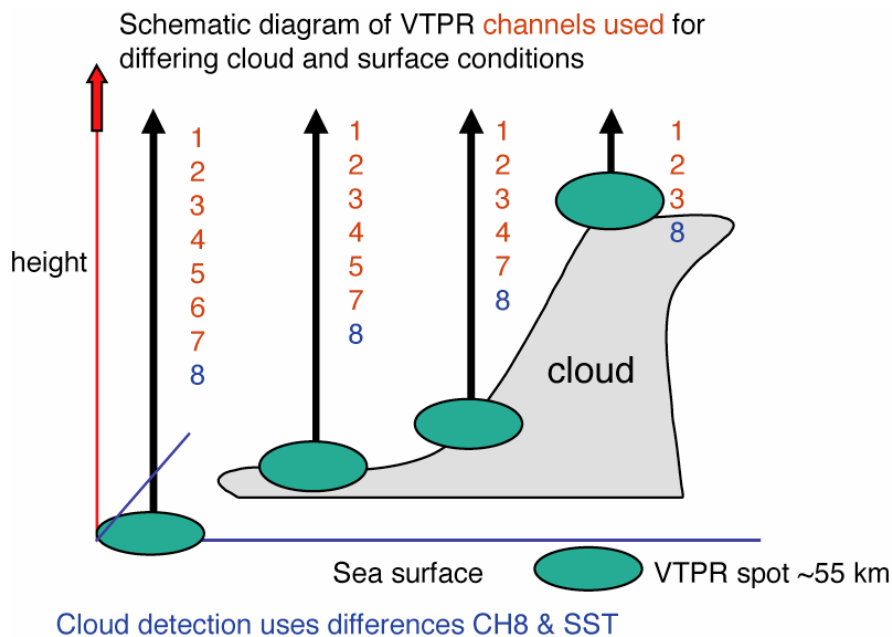


Figure 5 VTPR Cloud Detection

Image created from VTPR CH2 for all orbits on 28/12/1972

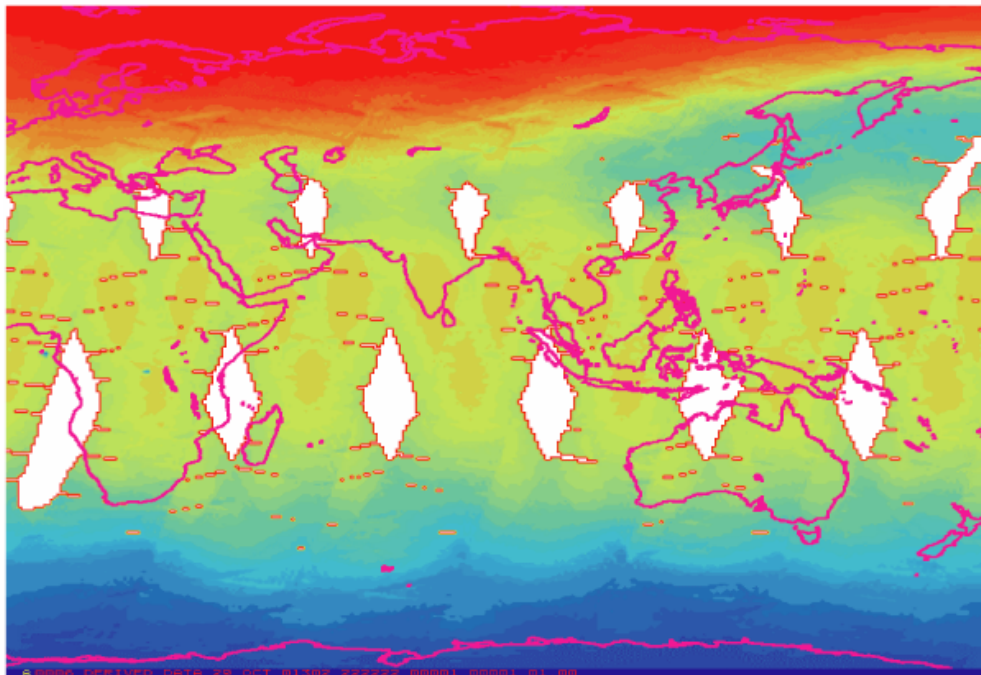


Figure 6 Raw VTPR radiance data for channel-2

VTPR data usage for two Tropospheric channels plotted on imagery from CH8: magenta CH6, yellow addition CH5

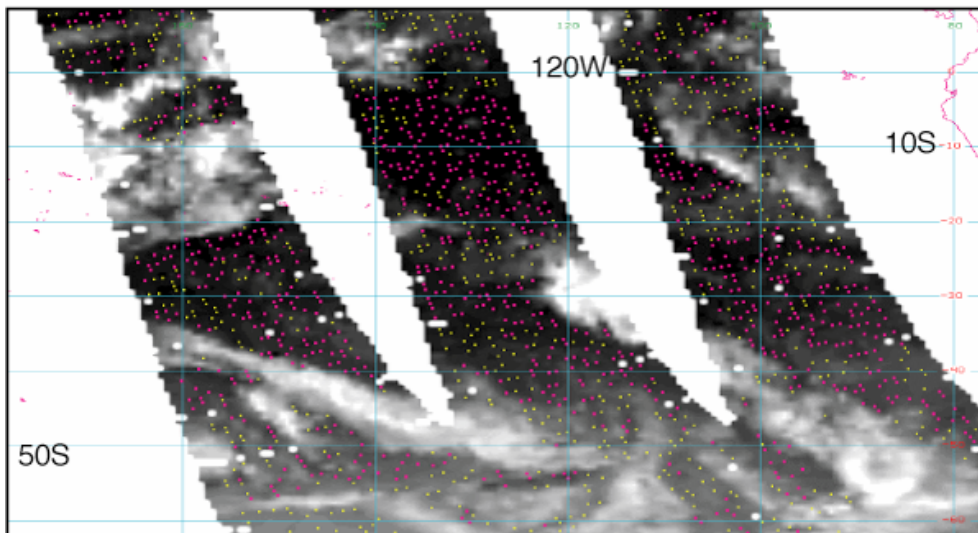


Figure 7 VTPR channels used in the analysis.

6. Scan and air mass bias corrections

The method used for the VTPR bias correction follows the method described by Harris and Kelly (2001). A scan bias and air mass correction are applied to the observed radiances. The scan bias correction is a function of the spot position in the scan line and the latitude. The scan bias and air mass corrections are calculated on an initial two week sample of each instrument. The air mass correction is a regression based method for each channel using the following predictors from the model first guess:

Thickness Z (1000-300) hPa

Thickness Z (200-50) hPa

Surface Skin Temperature

Total Column Water Vapour

Figure 8 shows examples of how much channels 1-7 have been adjusted (corrected) in reference to the first guess for both VTPR instruments 1 and 2 on NOAA-2. All curves represent statistics based on the observation-minus-first guess departures. The blue curves show the biases before bias correction and the solid red show the biases after the application of bias correction. The two dotted red lines show one standard deviation of the departures. Even though the standard deviations are about twice the values for the HIRS instrument, the use of VTPR radiances enabled good quality analyses over oceanic regions.

The analysis and first guess departures for VTPR channels 1-6 are shown in Figure 9. In a single six hour analysis cycle about 31000 channel-1 radiances were used. This number reduces to about 11000 radiances for the lower channel-3 and to about 5000 for the lowest peaking channel-6. After analysis the standard deviation of the radiance departures reduces to about half that of the first guess.

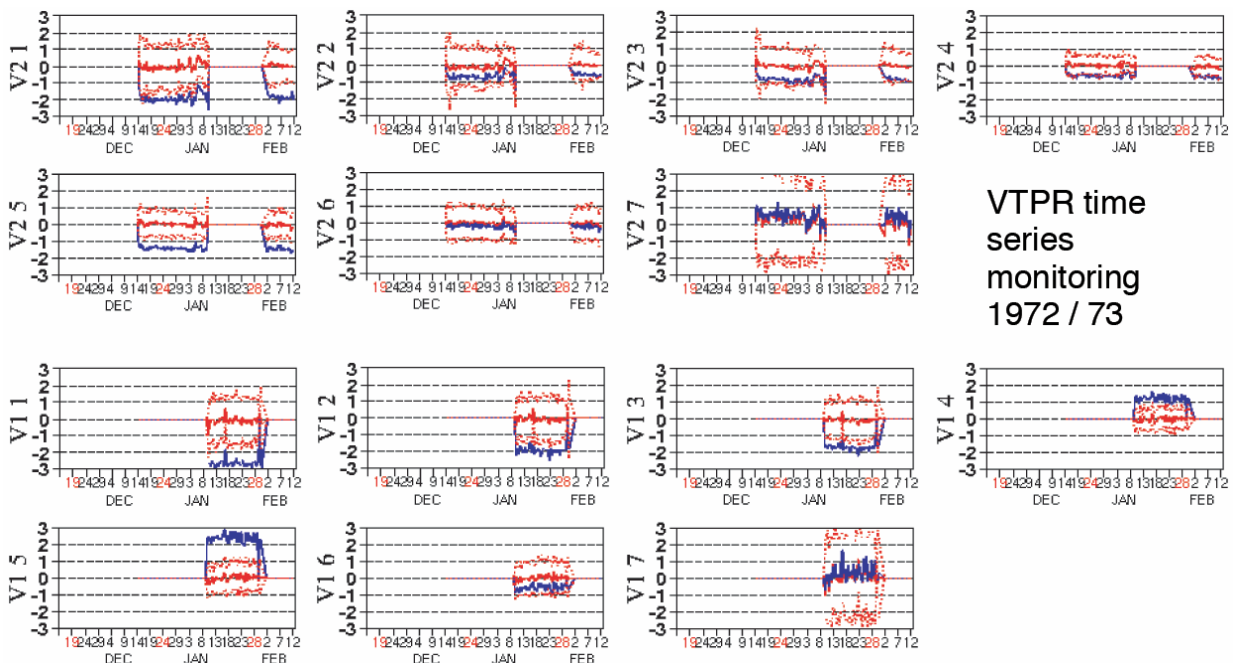


Figure 8 Bias correction time series for the two NOAA-2 VTPR sensors

Departure statistics for VTPR channels 1 to 6

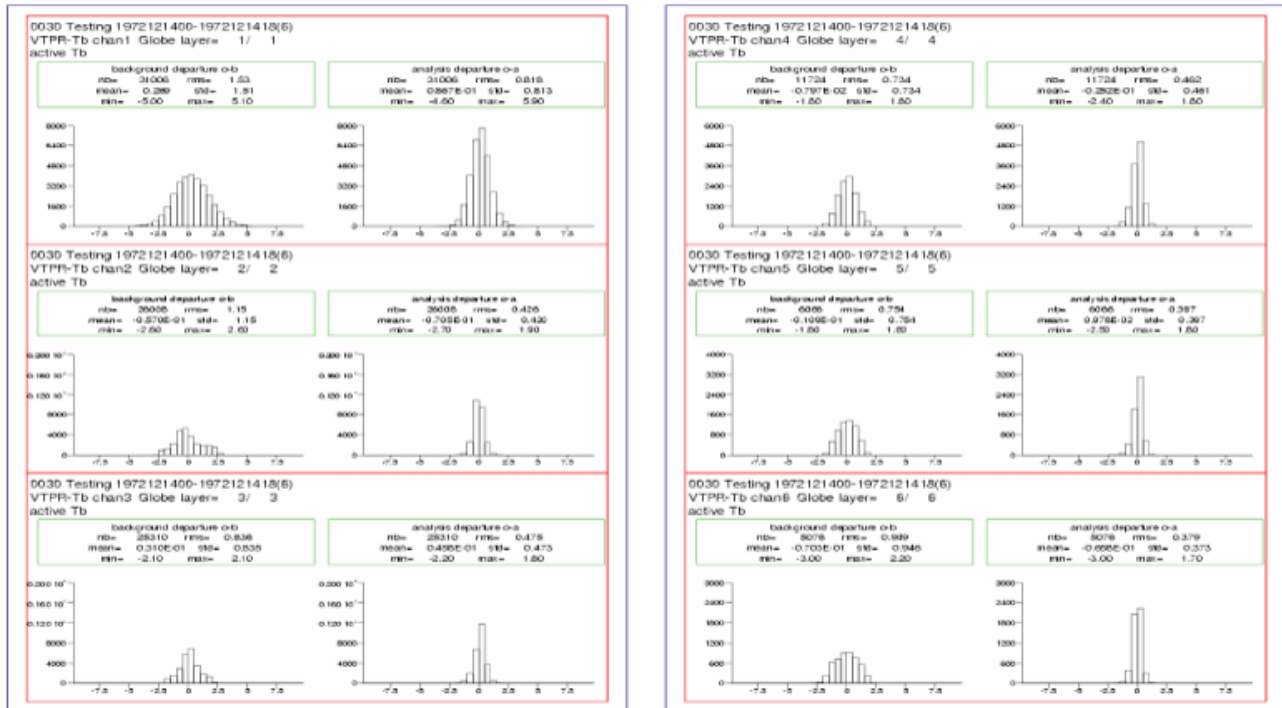


Figure 9 Departures of the bias corrected VTPR brightness temperatures from the analysis and first guess

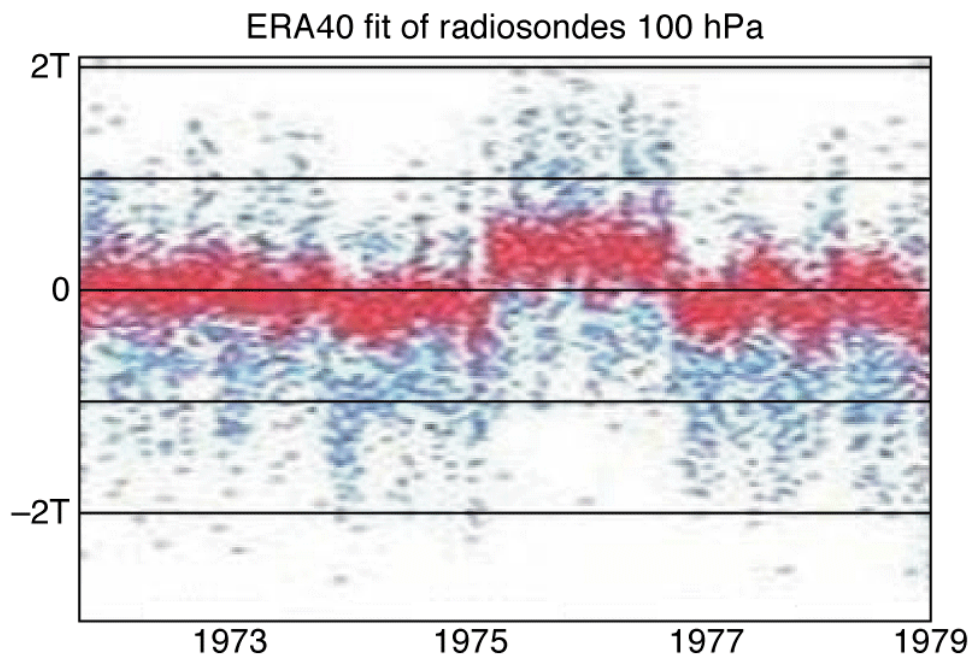


Figure 10 ERA-40 radiosonde departures at 100hPa.

Analysis increments of 500 hPa temperature and total precipitable water prior and during the introduction of VTPR radiances are shown in Figures 11 and 12. There is an increase in variance of temperature but the mean remains largely unchanged, whereas VTPR radiances make changes to the precipitable water. Periods of missing data in early January 1973 can be clearly identified due to the lack of variance in both figures.

From December 1974 until August 1976 the bias correction for NOAA 4 was incorrect resulting in some stratospheric temperature biases. Figure 10 shows the jump in the fit of ERA-40 to radiosondes in this period. ERA-40 analyses above 200 hPa should be treated with caution during this time.

Analysis Increments at the start of VTPR radiance assimilation

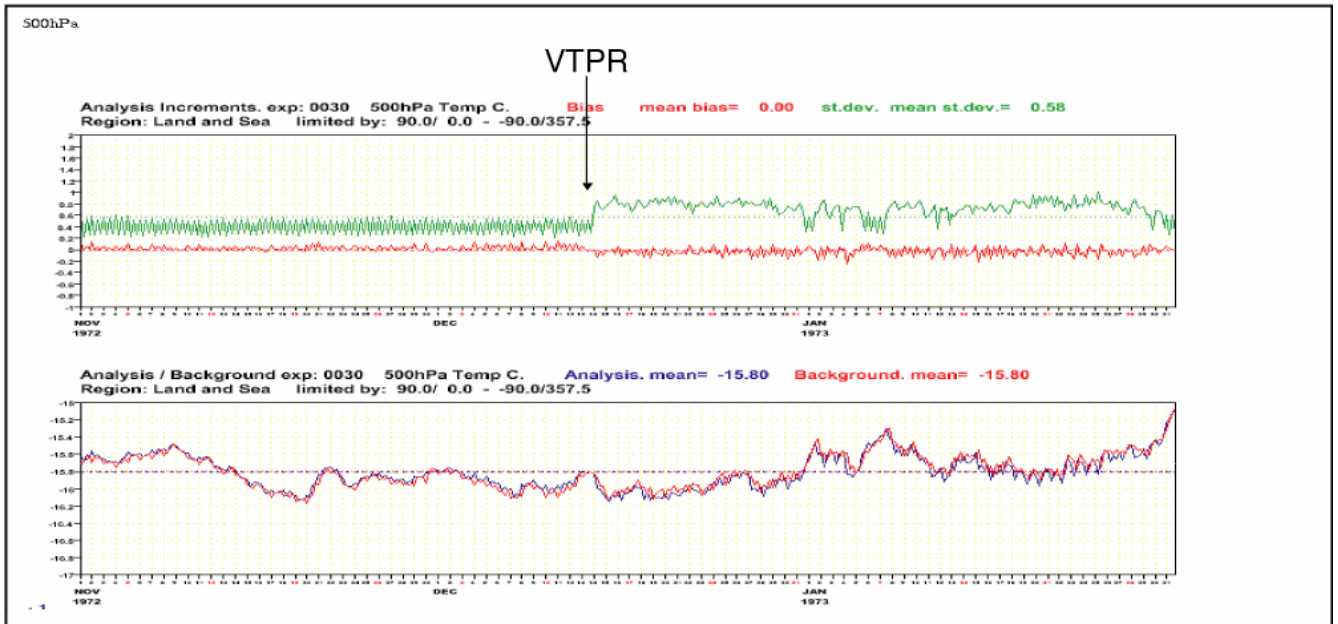


Figure 11 Time evolution of analysis increments of 500hPa temperature and the mean 500hPa temperature for both the background and analysis.

Analysis Increments at the start of VTPR radiance assimilation

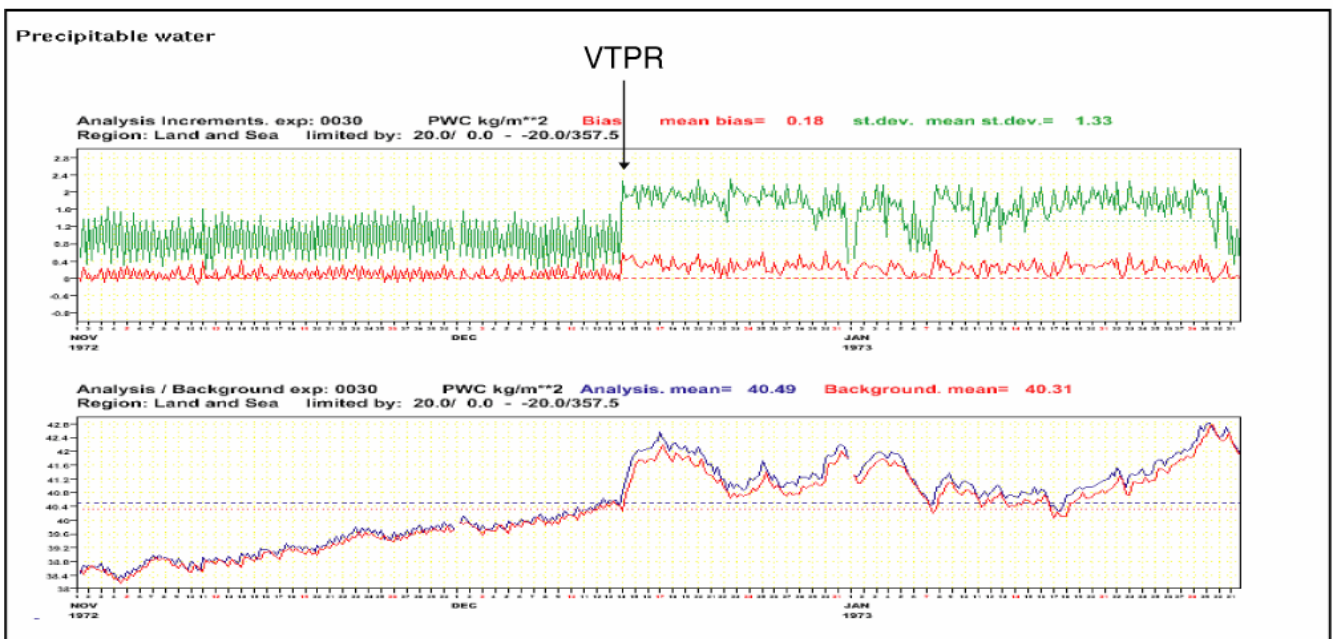


Figure 12 Time evolution of total precipitable water increments and mean values for both the background and analysis

7. Comparison of ERA-40 and the old Australian analysis

In 1974 a VTPR retrieval method was developed at ANMRC in Australia. It was based on a regression model and produced retrieved vertical profiles of temperature and moisture. These profiles were input to the Australian Hemispheric assimilation system (Kelly 1976).

It is interesting to compare the VTPR analysis from the Kelly (1976) assimilation with the same case from ERA-40 which used raw VTPR radiances. The 1000/500 hPa geopotential thickness for both analyses are shown in Figure 13. There is a clear similarity between these two analyses.

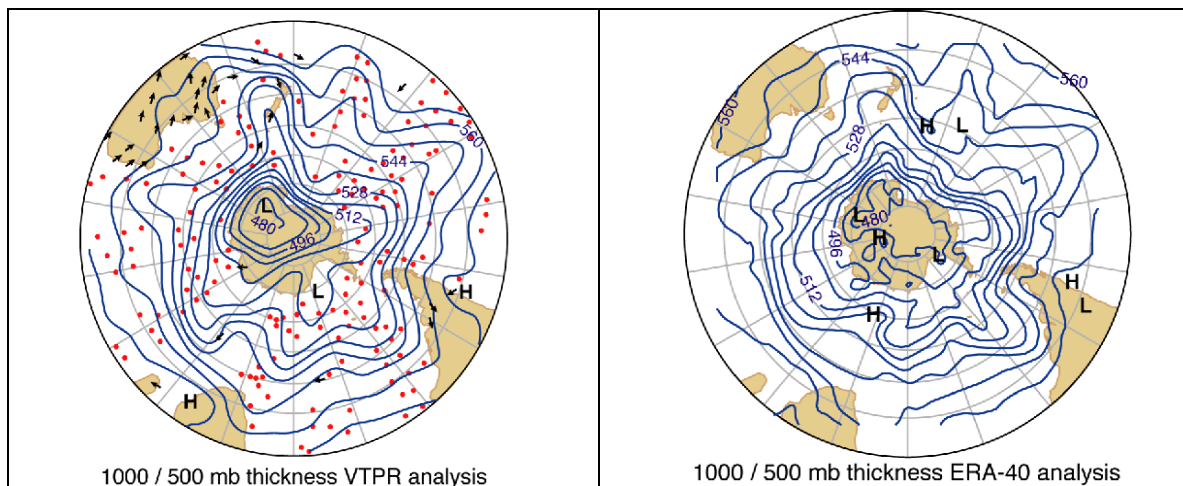


Figure 13 Comparison of ERA-40 with Australian analysis of VTPR.

8. Use of Australian PAOBs.

During the VTPR period (1973-1978) the surface data coverage in the Southern Hemisphere was poor. At this time only a few drifting buoys were active. The major source of surface oceanic data was cloud imagery.

In 1972 the Bureau of Meteorology in Australia introduced an operational numerical Southern Hemispheric analysis. At this time much of the source of data over the oceanic regions was satellite cloud imagery (Kelly, 1978). Using these images and all other sources of meteorological conventional data, manual analyses of mean sea level pressure were constructed. Next a grid of point values (known as PAOBs) were read off this manual analysis and were used as input to the Southern Hemispheric numerical ANalysis (SHAN). The method is described in a paper by Seaman and Hart (2003).

Unfortunately the PAOBs for this period were lost but the GRIB SHAN analyses were obtained from the Bureau Of Meteorology (BOM). PAOBs were then reconstructed by interpolation of the SHAN mean sea level fields on a regular latitude/longitude grid, see Figure 14. This grid was then used as PAOBs and used as part of the input observations for ERA-40 for the VTPR period.

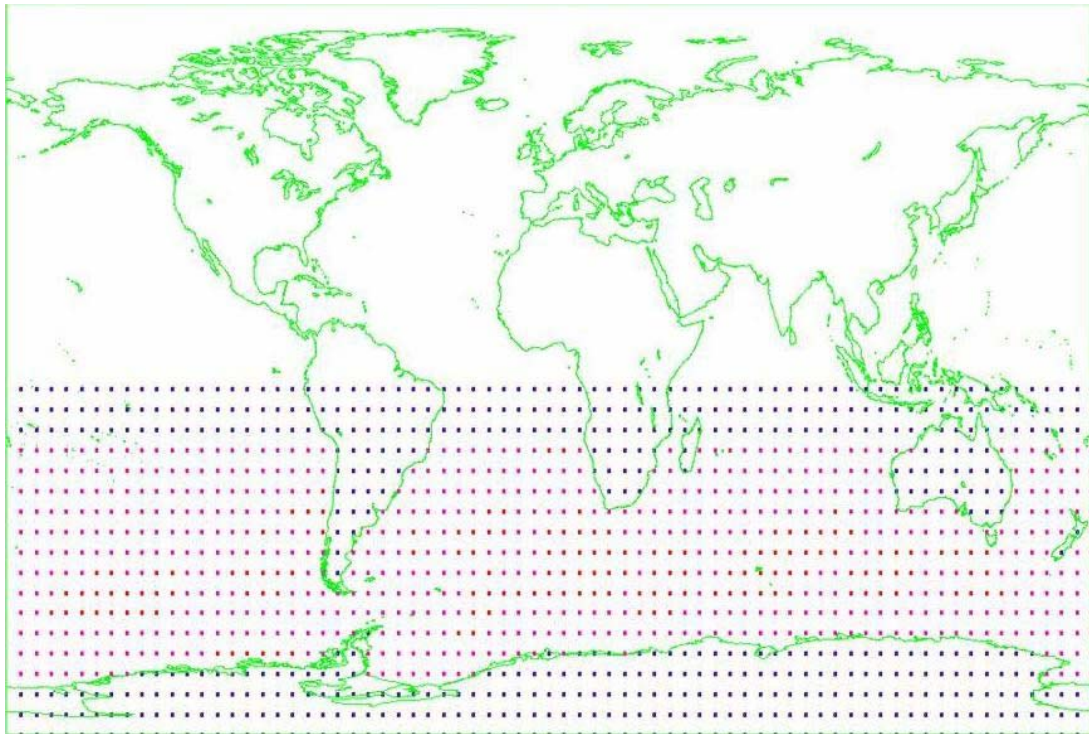


Figure 14 Data coverage of PAOBs (red data used in ERA-40)

9. Conclusions

ERA-40 has been a pioneering project in the use of the VTPR raw radiance data. After 25 years from the observation time the analysis system can now make optimal use of these valuable radiance data. A considerable effort was required to ‘clean up’ and earth-locate the VTPR raw radiance dataset.

Even if more information can be extracted from VTPR data in the future reanalyses ERA-40 results have already proven the positive impacts of VTPR data, see Uppala et. al (2005). Among those are the improved quality of TCWV over tropical oceans, better identification of tropical cyclones and also improved forecasts scores.

10. Acknowledgements

We would like to thank Roy Jenne and his group in NCAR to have had the foresight to keep the raw VTPR data in readable form through the years when there was no indication that these data would have such an importance in climate research.

11. References

- Hernandez, A., Kelly, G.A. and Uppala, S.M. 2004: The TOVS/ ATOVS observing system in ERA-40. ECMWF ERA-40 Project Report Series, 16, 49pp (available from www.ecmwf.int/publications).
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, A. Leetmaa, R. Reynolds, and R. Jenne. 1996: The NCEP/ NCAR Reanalysis Project. Bull. Amer. Meteor. Soc., 77, 437-471.

Kelly, G.A.M., P.E. Powers and F.J. Gauntlett. 1976: Temperature and water vapour retrievals from NOAA 4 satellite in the southern hemisphere. Proceeding of the Symposium on Meteorological Observations from Space: Their Contribution to the First GARP Global Experiment, June 8-10, 1976, Philadelphia, Penna.,USA, COSPAR, pp 77-84.

Kelly, G.A.M. 1978 Interpretation of satellite cloud mosaics for the southern hemisphere analysis and reference level specification. *Mon. Wea. Rev.*,106: 870-889.

McMillin L.M. et al, 1973: Satellite Infrared Soundings from NOAA Spacecraft. NOAA Technical Report NESS 65.

Saunders R.W., M. Matricardi and P.Brunel 1999 An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quart. J. Roy. Meteorol. Soc.* 125 pp 1407-1426.

Seaman R and Hart T, 2003. The history of PAOBs in the Bureau of Meteorology. *Aust Met Mag* 52,241-250.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. 2005: The ERA-40 Reanalysis. *Quart. J. Roy. Meteorol. Soc.*, in press.



Appendix A: Pre-processing the VTPR data set

The VTPR data available to ERA-40 are coded as Cray data (64 bit) sets. They are written sequentially in file0, file1,..., file44, orbit by orbit. Each orbit of data is stored as three separate files, corresponding to raw radiance (archive 1), cloud clear radiance (archive 2) and retrievals (archive 3). The coded data are written and therefore read one block at a time. The information necessary for decoding and obtaining a record (header, data or end) may come from a single block or multiple blocks. In the latter case, it means that the record is spanning blocks. For VTPR data, only the data records of archive 1 and archive 3 span blocks. Tables 3 and 4 show the format descriptions of the header record and data record of archive 1.

Table 3. The VTPR Archive-1 Format and Data Description:
Header Record (6480 bits)

Description	Number of words	Bits / word	Number of bits
Satellite ID	1	36	36
Start date(Year, Month, Day)	3	12	36
Day number in a year	1	12	12
Start time (Hour, Minute,Second)	3	12	36
End time (Hour, Minute,Second)	3	12	36
Calibration ID	1	12	12
Standard deviations	16	12	192
Fill	1	36	36
Orbit parameters	6	36	216
Epoch Date (Year, Month, Day)	3	12	36
Epoch time (Hour, Minute, Second)	3	12	36
Maximum Roll in the orbit	1	12	12
Right ascension of maximum Yaw	1	12	12
Fill	481	12	5772

Table 4. VTPR archive-1 format and data description:
Data record (19440 bits)
(one data record is a group of 3 records with 6480 bits each)

Description	Number of words	Bits / word	Number of bits
Radiances ($8 \times 23 \times 8$)	1472	12	17664
Latitude of spot (4,5)	1	12	12
Longitude of spot (4,5)	1	12	12
Latitude of spot (4,12)	1	12	12
Longitude of spot (4,12)	1	12	12
Latitude of spot (4,19)	1	12	12
Longitude of spot (4,19)	1	12	12
Hour, Minute, Second of spot (4,11)	3	12	36
Zenith Angles of 7 spots	7	12	84
Fill	1	12	12
Line count	1	12	12
Sea surface temperatures	184	8	1472
Fill	1	28	28
Instrument number	8	3	24
Clock for spot (1,1)	1	36	36

Processing program

A program to read and decode the VTPR data is available from NCAR (PROG1). Figure 15 shows the flowchart:

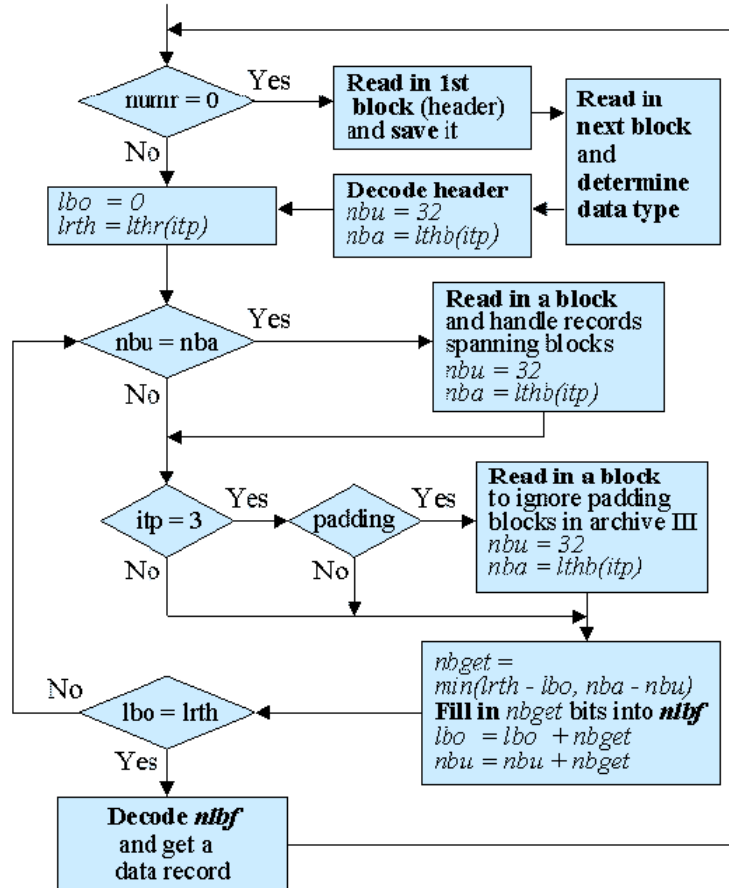


Figure 15 Flowchart of processing one file of VTPR data in original program.

1. Read in the first block of a file, regard it automatically as the header of the current file and save it for the decoding of the header record after the type index (itp) of the current file is determined.
2. Read in the next block, regard it automatically as a data record, determine itp , based on the length of this block (lth bytes).
3. Decode the header record.
4. Fill in the array $nlpf$, including reading in additional data blocks to handle the data records spanning blocks and ignore the padding blocks.
5. Decode $nlpf$ and get a data record when $lbo = lrth$.
6. When an end record is read in, set $numr = 0$ for processing the next file.

The meaning of the variables and arrays used in flowcharts and rejection criteria are given in Table 5.



Table 5. Variables and arrays used in flowcharts and rejection criteria

Parameter	Description
<i>nbl</i>	Array to store one block of data newly read in
<i>nbf</i>	Array to store all the data necessary for decoding
<i>nf</i>	File number
<i>nr</i>	Data record number of the current file
<i>numb</i>	Number of blocks already read in the current file
<i>itp</i>	Type ID (= 1, 2, 3) of the block newly read in
<i>itpc</i>	Type ID (= 1, 2, 3) of the block read in last time
$lrth = lthr(itp)$	Length of a data record in bits (= 19440, 720, 3280)
$lthb(itp)$	Length of a block in bits (= 6480, 720, 8160)
<i>nba</i>	$nba = lthb(itp) + 32$
<i>nbu</i>	Number of bits have been extracted in <i>nbl</i>
<i>nbget</i>	Number of bits extracted from <i>nbl</i> (filling in <i>nbf</i>)
<i>lbo</i>	Number of bits have been filled in <i>nbf</i>
$lthb(itp)$	Length of a block in bytes (= 816,96,1024)
<i>ltht</i>	Sum of the lengths of blocks ever read in for a data record. Usually, $ltht = 2446$ for $itp = 1$

There are three basic steps to obtain a data record:

1. Read the coded data blocks into array *nbl*.
2. Extract data from *nbl* and then fill them into *nbf* until it contains all the information from a record.
3. Decode *nbf* and get a data record.

Figure 16 demonstrates the procedure to get a VTPR-1c data record.

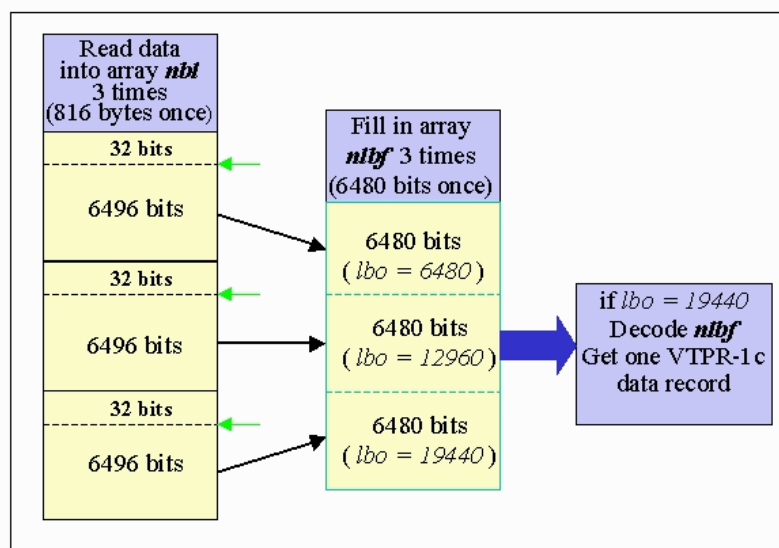


Figure 16 Flowchart of processing one file of VTPR data in the original program.



Establishment of rejection criteria for a data record

To guarantee data quality, rejection criteria must be set up and applied to the full data set. This is especially necessary when the quality control in the data assimilation system is unable to identify the erroneous data. This needs a detailed inventory, i.e., the full data set is read by the processing program. Every element of a data record, especially of the raw radiance data records, is saved and then checked carefully.

Based on the inventories, the rejection criteria have been set up as follows:

1. Criterion 1: The return information from the reading of a data record is 2 (error).
2. Criterion 2: The `ltht` (see 4.5) is not equal to 2448.
3. Criterion 3: The time (hour, minute, second) of a data record is unreasonable.
4. Criterion 4: The starting time (year, month, day, hour, minute, second) of an orbit is unreasonable.
5. Criterion 5: The time of a data record is earlier than the starting time of the current orbit.

Here, "unreasonable" means that the value is beyond the normal range. Only the time parameters of the elements of a data record are used to set up the criteria. In practice, criterion 3 is important for archive 1 data. Since the time of a data record is in the third part of `n1bf`, if it is right, usually everything is right. The exception is when criterion 2 applies.

These five rejections happen independently, therefore, if `err`, `rej1`, `rej2`, `rej3`, `rej4` is the number of the data records rejected by criterion 1, 2, 3, 4, 5 respectively, the total number of the rejected data records is:

$$\text{errt} = \text{err} + \text{rej1} + \text{rej2} + \text{rej3} + \text{rej4}.$$

If `nr` is the total number of records, then the error rate is:

$$\text{err_rate} = \text{errt}/\text{nr}$$

If the rejection criteria are applied, how many data records will be rejected? To answer this question, `Inventory_1` is performed. This can help us understand the quality of the whole of the VTPR data. It also provides a verification of the rejection criteria themselves.

Inventory_1:

- Full VTPR data set
- `PROG1` + Rejection criteria (1~5)

A careful check indicates that the rejection criteria work very well. Therefore, the rejection rate is really the error rate. From Table 6, the average error rate of the VTPR-1c data is 2.95%.



Table 6. Numbers and error rates of the VTPR-1c data records in file0 ~ file44: Results from Inventory 1.

file	err	rej1	rej2	rej3	rej4	errt	nr	err_rate
0	86	146	2601	594	34	3461	40959	8.45%
1	65	517	3106	166	8	3862	37827	10.21%
2	30	208	1526	132	28	1924	37612	5.12%
3	49	68	2322	225	5	2669	38281	6.97%
4	34	56	1523	42	16	1671	37109	4.50%
5	42	46	2151	21	37	2297	26153	8.78%
6	50	41	2375	100	33	2599	33679	7.72%
7	50	10	2517	484	24	3085	38928	7.92%
8	74	129	4767	372	9	5351	37161	14.40%
9	19	25	1091	81	0	1216	35800	3.40%
10	3	46	144	1	29	223	37825	0.59%
11	0	10	14	24	0	48	25545	0.19%
12	0	8	0	0	0	8	17879	0.04%
13	0	84	0	0	51	135	37748	0.36%
14	0	16	3	0	0	19	41890	0.05%
15	2	7	0	0	0	9	40617	0.02%
16	1	15	23	51	0	90	38298	0.23%
17	0	7	0	0	0	7	41332	0.02%
18	0	22	0	0	0	22	39989	0.06%
19	0	178	0	0	0	178	39967	0.45%
20	0	10	0	0	0	10	38895	0.03%
21	0	4	5	0	0	9	33520	0.03%
22	0	42	637	0	0	679	41397	1.64%
23	0	27	3090	0	36	3153	42365	7.44%
24	5	3	1402	0	38	1448	42815	3.38%
25	0	27	1254	0	1	1282	42920	2.99%
26	1	202	61	167	36	467	40532	1.15%
27	0	16	0	0	106	122	43030	0.28%
28	2	7	220	0	0	229	41310	0.55%
29	2	417	76	63	0	558	42865	1.30%
30	0	142	0	0	74	216	41931	0.52%
31	2	193	118	0	0	313	40416	0.77%
32	0	69	0	0	0	69	27326	0.25%
33	1	87	58	0	36	182	42968	0.42%
34	2	546	104	0	36	688	43304	1.59%
35	1	487	124	0	0	612	42989	1.42%
36	10	327	510	0	0	847	41431	2.04%
37	2	410	438	0	0	850	43213	1.97%
38	2	805	104	0	0	911	43130	2.11%
39	6	3863	82	13	150	4114	42121	9.77%
40	4	3053	322	9	36	3424	39435	8.68%
41	0	171	20	0	0	191	37673	0.51%
42	0	325	1	0	0	326	39312	0.83%
43	0	218	90	63	0	371	31032	1.20%
44	0	0	0	54	0	54	1586	3.40%
Total	545	13090	32879	2662	823	49999	1692115	2.95%

Basically, there are no great problems. The errors happened mainly in three periods:

- i) November 1972 <96> April 1974 (file0 - file9), this corresponds to the first period when NOAA-2 data was available.
- ii) January 1976 <96> August 1976 (file2 - file25), this corresponds to the end of NOAA-4 data.
- iii) June 1978 <96> October 1978 (file39 - file40), this is in the period of NOAA-5.

In these 3 periods, the error rates are not negligible and can very often reach about 8%.



Modifying the processing program and data recovery

All the information of the VTPR data should be contained in the original VTPR data set. This means that it should be possible to recover the data through modifying the processing program. The reasons for the erroneous VTPR data have been revealed by the more complete inventories, e.g., the check of the length of every block. Generally, the reason is that PROG1 is unable to handle some unexpected situations in the VTPR data. Although every block in the VTPR data set is correctly formulated, and both step 1 (reading in blocks) and step 3 (decoding) (see Figure 4) are always realised correctly in PROG1, unfortunately, step 2 (extracting and filling in) occasionally does not interpret the data correctly. This leads to erroneously decoded header or data records.

One example of an unexpected situation is when there is no end record in a data file. If file-1 is an archive 3 file ($itp=3$) without an end record, the next file (file-2), which is, for instance, an archive 1 file ($itp=1$), will be incorrectly categorized as an archive 3 file by PROG1. This happens because, in PROG1, itp is assessed and then updated only if the first block of a file is read in. The decoding routines depend on itp , so when itp is wrong, the decoded record is wrong.

Similarly, if there is no header record in file-2, the first block of file-2, which is actually a data block, will be regarded as a header record. In order to handle such unexpected situations, PROG1 has been modified and a new processing program, PROG2, developed.

PROG2 is based on the concept that data handing at block level is sufficient for all processes, including all the 3 steps in Figure 17. This make PROG2 be able to handle the situation in which a file is not complete (no end record or header record).

Figure 5 is the flowchart of PROG2. The main modifications in PROG2 include:

1. Additional assessment is made to determine whether a block is a header record or a data record.
2. The block type is assessed every time a new block is read in, rather than only when the first block of a file is read in as PROG1.
3. For archive 1 data, when a decoded data record is rejected by criterion 3, one data block is skipped.

Inventory_2 was undertaken to evaluate the effects of PROG2.

- Full VTPR data set
- PROG_1 + Rejection criteria (1-5)

From Table 7 and Figure 18, we can see that most of the erroneous VTPR-1c data have been recovered. The number of erroneous VTPR-1c data records has been reduced from 49999, corresponding to 58 days' data, to 14961, corresponding to 17 days' data. The average error rate has reduced from 2.95% to 0.88%. The three periods with a large error rate have been reduced to one period.

An unsolved problem in data recovery

The error rates in Inventory_2 are still about 9% in file39 and file40. The reason is that the length of many archive-1 blocks in these two files is longer than 816 bytes (normal), for instance, 820, 824, 828 and so on, and $ltht$ is larger than 2446 bytes, so rejection criterion 2 occurs very often. In fact, this means that there is something wrong with the data block. PROG2 cannot deal with this problem.

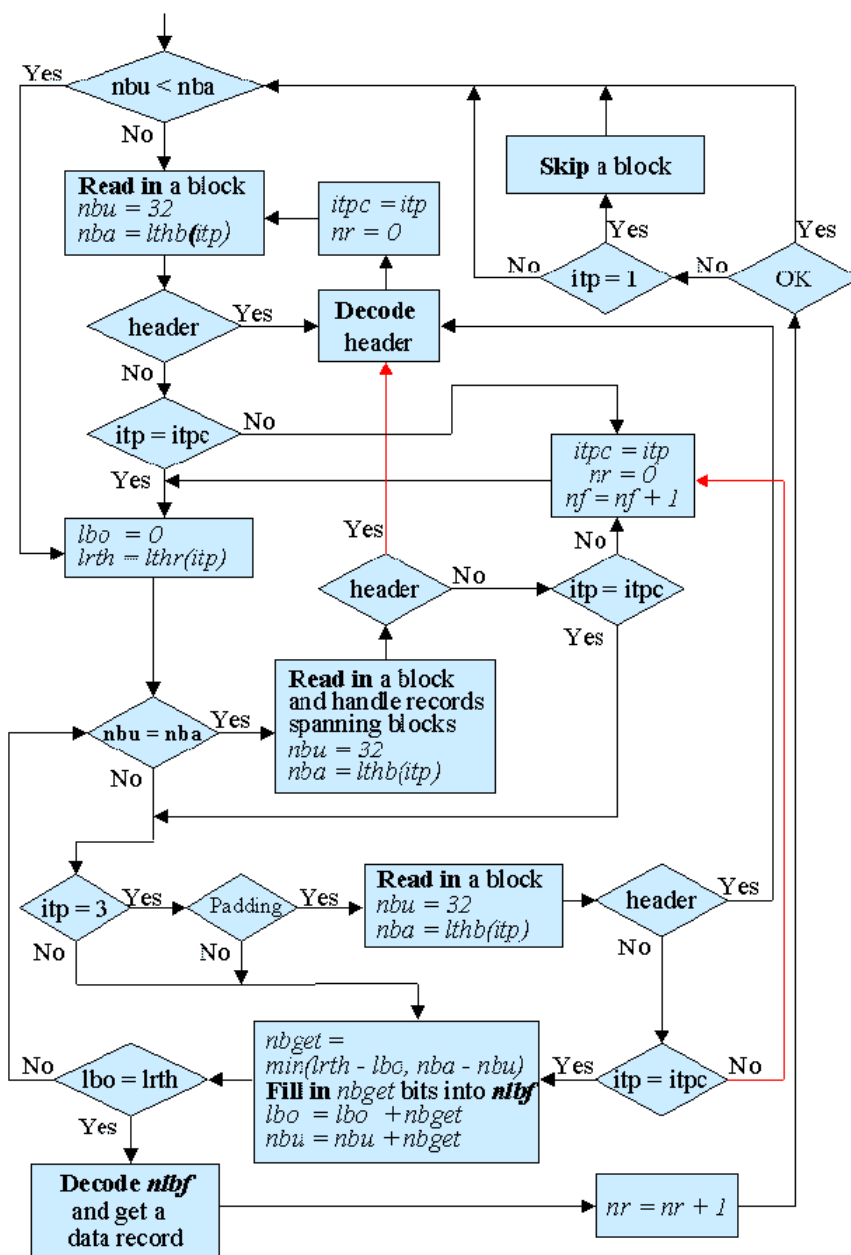


Figure 17 Flowchart of VTPR data processing in the modified program.



Table 7. Numbers and error rates of the VTPR-1c data records in file0 ~ file44: Results from Inventory 2.

file	err	rej1	rej2	rej3	rej4	errt	nr	err_rate
0	0	16	70	0	34	120	40938	0.29%
1	0	8	163	0	8	179	37489	0.48%
2	0	6	44	0	0	50	37438	0.13%
3	0	27	182	58	5	272	38209	0.71%
4	0	4	100	0	16	120	37171	0.32%
5	0	2	87	0	37	126	26065	0.48%
6	0	13	113	65	33	224	33662	0.67%
7	0	12	115	0	24	151	38998	0.39%
8	0	22	408	0	8	438	37176	1.18%
9	0	24	137	0	0	161	35758	0.45%
10	0	47	5	0	30	82	37820	0.22%
11	0	10	1	0	0	11	25545	0.04%
12	0	80	2	0	0	82	34000	0.24%
13	0	84	0	0	51	135	37748	0.36%
14	0	16	9	0	0	25	41887	0.06%
15	0	7	0	0	0	7	40615	0.02%
16	0	15	8	0	0	23	38294	0.06%
17	0	7	0	0	0	7	41332	0.02%
18	0	21	0	0	0	21	39988	0.05%
19	0	4	0	0	0	4	39793	0.01%
20	0	9	0	0	0	9	38894	0.02%
21	0	4	14	0	0	18	33515	0.05%
22	0	42	49	0	0	91	41373	0.22%
23	0	24	168	0	36	228	42280	0.54%
24	0	0	80	0	38	118	42775	0.28%
25	0	26	81	0	1	108	42882	0.25%
26	0	197	24	41	36	298	40520	0.74%
27	0	16	0	0	106	122	43030	0.28%
28	0	8	2	0	0	10	41309	0.02%
29	0	406	7	0	0	413	42852	0.96%
30	0	139	7	0	74	220	41926	0.52%
31	0	191	2	0	0	193	40412	0.48%
32	0	65	0	0	0	65	27322	0.24%
33	0	84	4	0	36	124	42963	0.29%
34	0	554	3	0	36	593	43309	1.37%
35	0	484	3	0	0	487	42985	1.13%
36	0	330	8	0	0	338	41425	0.82%
37	0	393	96	0	0	489	43162	1.13%
38	0	806	3	0	0	809	43131	1.88%
39	0	3867	17	0	150	4034	42122	9.58%
40	0	3055	21	0	36	3112	39432	7.89%
41	0	170	4	0	0	174	37671	0.46%
42	0	325	5	0	0	330	39310	0.84%
43	0	218	8	60	0	286	31028	0.92%
44	0	0	1	53	0	54	1586	3.40%
Total	0	11838	2051	277	795	14961	1707140	0.88%

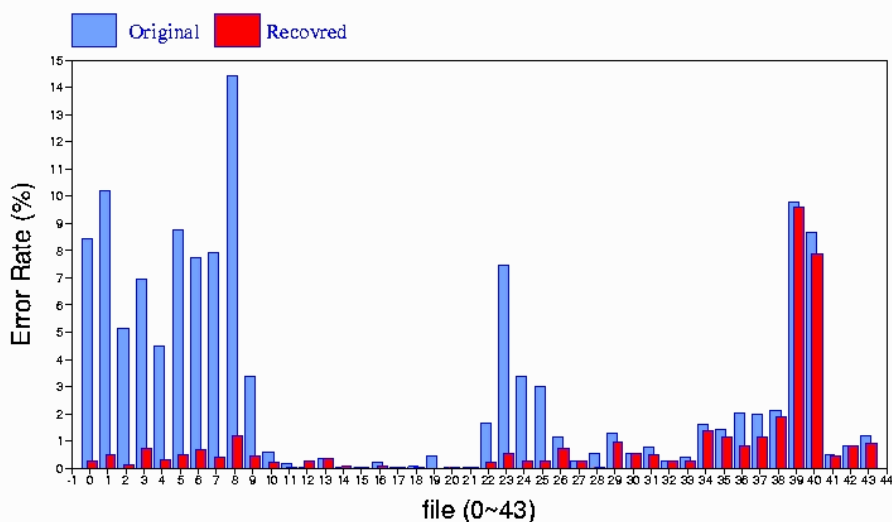


Figure 18 The error rates of the VTPR-1c data in file0 -file43.