Automatic data processing and quality control: experiences from ISO-LWS

M. J. Burgdorf[a], A. S. Harwood[b], N. R. Trams[a], T. L. Lim[a, c],
S. D. Sidher[b], B. M. Swinyard[b] and P. E. Clegg[c]

[a]ISO Science Operations Centre, ESA Science Department,
Villafranca del Castillo, P.O. Box Apdo 50727, 28080 Madrid, Spain

[b]Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK

[c]Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, UK

ABSTRACT

The high level of automation in the operation of the ESA Infrared Space Observatory, together with high observing efficiency, leads to a requirement for a commensurate level of automation in the subsequent processing of the astronomical data. This inevitably means that all data for a given instrument mode have the same calibration applied, regardless of the exact details of the object being observed. Questions then arise about these "pipeline processed" data in terms of the calibration accuracy achieved; how to control the quality of data received by the observer and how much further processing is required - or desirable - by the observer. In this paper we outline the experience of two years of operation of the Long Wavelength Spectrometer on board ISO, detailing the improvements made in the pipeline processing during this time and the difficulties encountered in the automated processing of some instrument modes.

Keywords: Data processing, observatories

1. INTRODUCTION

The Long Wavelength Spectrometer[1] is one of two complementary spectrometers aboard the European Space Agency's Infrared Space Observatory.[2] It covers the wavelength range 43 - 197 μm at either medium or high spectral resolving power. There are ten doped-Germanium photoconductors, all of them being used simultaneously in different wavelength ranges such that a full grating spectrum is divided into ten subspectra.

Like the other instruments on ISO, the LWS is operated in a pre-scheduled way in order to maximize the efficiency of this observatory. This means that the details of the observations have to be defined well in advance. Four different Astronomical Observation Templates (AOT's) are available to users of the LWS for entering their observations; they correspond to the basic instrument modes: scans over a line or an extended wavelength range with the grating or the Fabry-Perot subsystem. By means of the Proposal Generation Aids software an astronomer assigns concrete values to the variables in these templates (coordinates, integration times etc.) and gets his observation into the "Mission Database". From the pool of observations of all instruments a timeline is built for each revolution of the satellite. The observations are then executed by translating them into instrument command sequences which are sent to the spacecraft. These sequences are always of the same structure for a given AOT.

The standardization of the astronomical observations makes an automatic processing of the generated data possible. One branch of this is the real-time monitoring of the instruments which guarantees that none of the housekeeping values goes out of limits unnoticed,[3] the other is the off-line processing which generates the data products that are sent to observers on CD-ROM. In section 2 we describe in detail the various steps of generating

Other author information: (Send correspondence to M.J.B.)
M.J.B.: E-mail: mburgdor@iso.vilspa.esa.es
A.S.H.: E-mail: ash@coyote.bnsc.rl.ac.uk
N.R.T.: E-mail: ntrams@iso.vilspa.esa.es
T.L.L.: E-mail: tlim@iso.vilspa.esa.es
S.D.S.: E-mail: sds@ast.star.rl.ac.uk
B.M.S.: E-mail: bms@coyote.bnsc.rl.ac.uk
P.E.C.: E-mail: P.E.Clegg@qmw.ac.uk

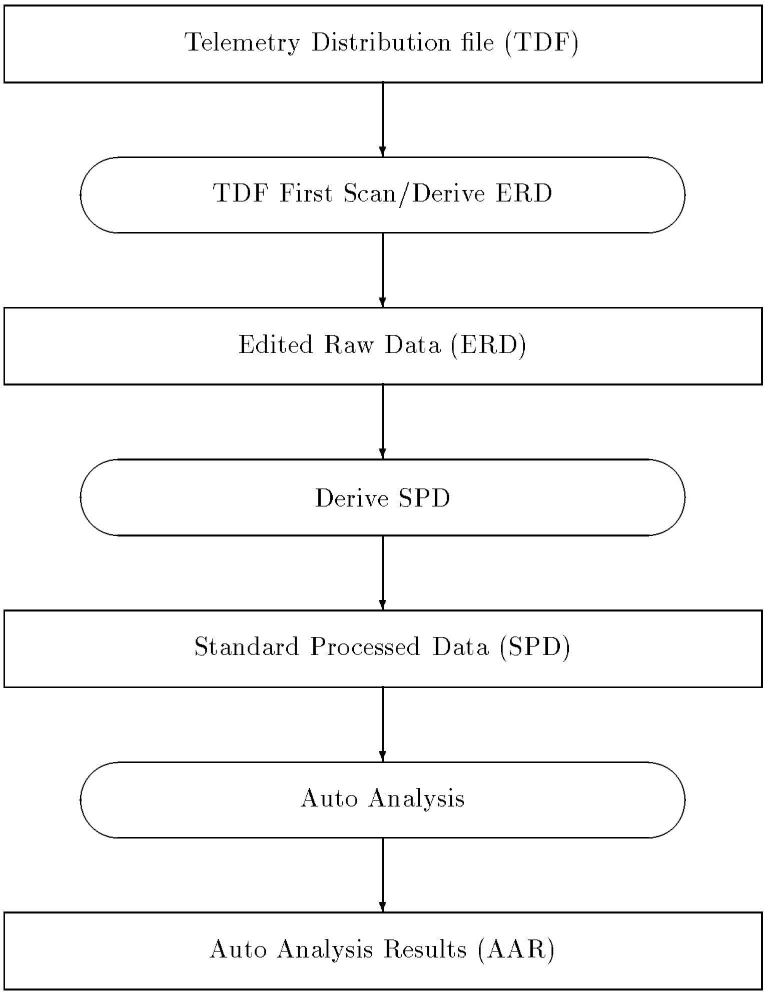

Figure 1. Overview of the ISO data processing

these "pipeline" products and in section 3 we discuss the modifications that became necessary from experience in flight. Subsequently section 4 gives an overview of the possibilities to examine the data products with LWS Interactive Analysis (LIA) and ISO Spectroscopy Analysis Package (ISAP) routines, and we finish this paper with a description of some instrument peculiarities which can currently not be handled, yet, by the automatic processing.

2. PIPELINE PROCESSING

The automated ISO data processing for each instrument consists of three stages (see figure 1) which are explained with the example of the LWS in the following subchapters. The products from each stage are sent to the observer so that he can choose at which level to start the analysis of his data.

2.1. Derivation of Edited Raw Data

The first step in the ISO processing is in principle the same for all instruments: the telemetry data from the satellite are related to the different observations executed. For this the start and the end of an observation and important events within like beginning and end of an illuminator flash are identified and written to a "Compact Status History" file. This information is then used to extract a set of Flexible Image Transport System (FITS) binary tables with Edited Raw Data. The part of the ERD that is common to all instruments includes housekeeping and satellite information and has the same format for each observation whereas the instrument specific ERD contain mainly the science data for a given observation. The ERD are a complete set of data in a primitive form, therefore they are not a useful starting point for further analysis by the observer.

2.2. Derivation of Standard Processed Data

The second stage of pipeline processing converts the raw data from the detectors into photocurrents. This is carried out via the following steps:

- Construction of ramps and discarding unusable readouts: The detector signals are grouped into the useful time intervals between resets of the integrating amplifiers. The periods where the instrument was changing between modes and therefore was not in a well-defined state are discarded.
- Conversion of read-outs to voltages: A linear relationship is used to convert the digital units of each elementary integration ramp into Volts; the coefficients needed for that operation are read from calibration files.
- Discard saturated readouts: All voltages above a certain threshold that is defined in a calibration file are discarded. The number of saturated readouts is recorded in the SPD.
- First level deglitching: The recognition of glitches, i. e. sudden jumps in output signal caused by hits of highly energetic particles, is done in several steps. First the differences between adjacent points are calculated and the mean and standard deviation of this set of data are determined. Those values in the set that are further away from the mean than 3.5 standard deviations are checked for the typical glitch pattern in the respective readouts. Once a candidate for a glitch event is identified, its height is calculated and compared to the overall signal in order to recognize insignificant outliers. The rest of an integration ramp after a (positive) glitch is discarded and also the two following ramps.
- Extraction of ramp slopes and uncertainties: A second order polynomial is fitted to each ramp and its slope is determined from the first and the last point on that fit. With a linear function the slopes are then converted into photocurrents; if a ramp has been discarded, e. g. because of a glitch, its value is set to zero.
- Illuminator processing: Every scientific observation contains measurements of the dark background and the signal from the internal illuminator flashes for reference. The calculation of the photocurrents is done for them in the same way as for the observation on the target except for a slight difference in the deglitching algorithm.

This processing is independent of the AOT used; its product contains scientific data, but not in astronomical units. Nevertheless it will be used by expert users as starting point for an interactive analysis of the data, if they want to use e. g. their own flux calibration.

2.3. Auto Analysis

Auto Analysis applies the flux and wavelength calibration to the SPD and creates a data product in scientifically meaningful units. It performs the following steps:

- Dark current / background straylight subtraction: The average "dark signal" as measured before and after each astronomical observation is subtracted from the signal of the source.
- Grating scan wavelength calibration: The positions of the grating mechanism are converted into wavelengths by using a table in a calibration file.

- Grating spectral responsivity calibration: As the instrumental responsivity varies as a function of wavelength, each flux value obtained in grating mode is corrected with the relative spectral response function of the instrument. This correction is done by dividing the fluxes by the responsivity values in a calibration table.

- Spectral bandwidth correction: For grating scans the flux of each detector is divided by the FWHM spectral resolution.

- Fabry-Perot (FP) scan wavelength calibration: In order to determine the wavelength for a Fabry-Perot scan, the mechanism position is converted into distance between the two FP etalons. The pipeline calculates then with this value and with the order in which the FP was operated the wavelength.

- Fabry-Perot spectral responsivity calibration: The fluxes in an FP spectrum are divided by the responsivity values for the respective FP and grating positions.

- Velocity correction to wavelength: The velocity of the spacecraft towards the target is included in the header of the SPD file; this information is used for a correction of the measured wavelength by the Doppler effect.

At this stage an intermediate Auto Analysis product is generated, because it is possible, especially for weak sources, that the following steps cause systematic errors. In these cases the observer should perform the remaining processing with interactive analysis (see section 4).

- Absolute responsivity correction: At the beginning and at the end of each scientific observation a sequence of standard illuminator flashes is performed. The ratio between the signal from a reference illuminator flash and the one from these sequences gives a correction factor with which the absolute responsivity for each detector is multiplied. As the measurement of the standard illuminator flash takes only 100 sec, it is prone to corruption e. g. by strong glitches and this can lead to an incorrect absolute responsivity correction.

- Responsivity drift correction: The wavelength range AOT's consist of several identical scans with the grating or the FP, and their average signal is used to correct for drifts within an observation. Sometimes the drift slope cannot be calculated because of a lack of valid data.

The AOT-specific Auto Analysis produces for the bulk of observations reliable spectra that give the flux in $W/cm^2/\mu m$ (W/cm^2 for FP spectra) as a function of wavelength in μm. It can serve as a good starting point for further work, e. g. removing remaining instrument peculiarities like fringes or extracting scientific results. If problems occur during the processing Auto Analysis will generate one out of 52 standard error messages. Typical examples are the notification of the last point in a scan having a different direction or of observations where the measurement of the dark current is doubtful, because too many elementary integration ramps had to be rejected due to glitches. These messages are brought to the attention of an instrument specialist and are checked interactively - mostly with the result that the data products are valid nevertheless.

3. IMPROVEMENTS OF THE PIPELINE DURING THE MISSION

The experience gained from analyzing pipeline products during the mission leads to constant improvements of calibration files and of the routines for the derivation of both SPD and AAR. At certain intervals new versions of the pipeline are generated which include these updates, and all observations obtained until then are reprocessed.

3.1. Modifications to Calibration Files

Improvements in the calibration of the instrument result usually only in more accurate numbers and do not require modifications of the processing routines. These values are embodied in updates of calibration files that are used in the pipeline processing. For the derivation of Standard Processed Data there are eleven such files which contain, in the form of FITS headers, the parameters needed to perform deglitching and voltage conversion. The Auto Analysis stage makes use of FITS binary tables or primary arrays that contain e. g. the relationship between mechanism position and wavelength or the relative response of the instrument.

The dramatic improvements in the quality of the data that were achieved by updates of calibration files are best illustrated with the example of the responsivity calibration. The version of the pipeline that was used immediately after launch still relied on measurements in the laboratory, where extended emission from a blackbody was used for

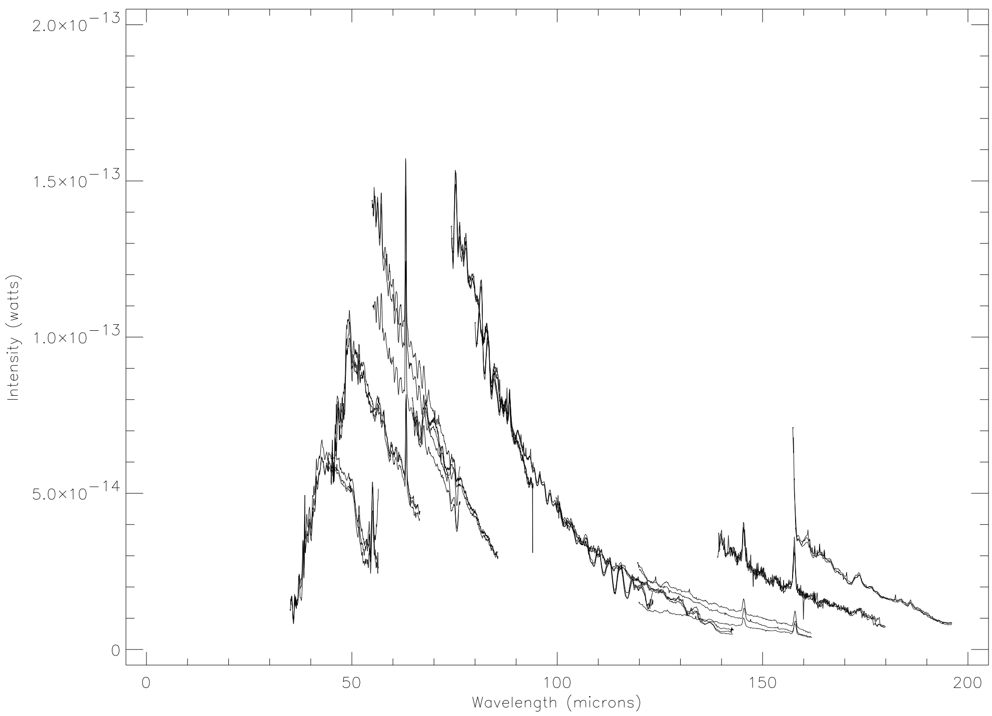

Figure 2. *The LWS end-to-end spectrum of the bright HII-region S106 as observed in the performance verification phase of the satellite and processed to Auto Analysis level with a pipeline version developed before launch. Because the responsivity of the detectors was very different in flight from the tests in the laboratory the ten subspectra do not match. (Another unit for intensity was used in later versions of the pipeline.)*

flux calibration. The resulting spectra of observations in flight had discontinuities because of the changed detector sensitivities and they showed fringes which had not been recognized in the relative spectral response function (RSRF, see figure 2). New values for the absolute and relative responsivity that were based on observations of Ceres and later of Uranus solved both problems largely (see figure 3).

3.2. Modifications to the Derivation of Standard Processed Data

Apart from several minor bug fixes and modifications to the layout and content of the SPD files the changes here applied mainly to the determination of the slope of an elementary integration ramp (see figures 4 and 5). One requirement for this is the correct identification of useless points, and it became necessary to increase the amount of time discarded at the beginning of a ramp at low signal levels, because it took the detectors longer than expected to recover from a reset pulse. The slope itself was at first determined by fitting a first or second order polynomial to the data, but this led at times to discontinuities in the spectra. Hence finally all ramps were fitted by a second order polynomial and the slope was determined from a straight line through the first and the last point of that fit.[4] This means that the result lies between the higher slope at the beginning and the lower one at the end of the ramp, if the photocurrent decreases because of saturation.

3.3. Modifications to the Derivation of Auto Analysis Results

Major changes to the Auto Analysis were the implementation of a time dependent grating wavelength calibration, the change in calculating the dark signal and the number of files to be produced. Most work was concentrated, however, on optimizing the drift correction. At first the use of illuminator flashes to correct for changes in responsivity within a revolution[5] produced unreliable results, therefore it was switched off. But a new illuminator flash sequence made this correction within the pipeline possible. Besides, an algorithm was implemented which removed drifts within an observation by calculating the average signal from each scan. The usually positive slope of these average detector

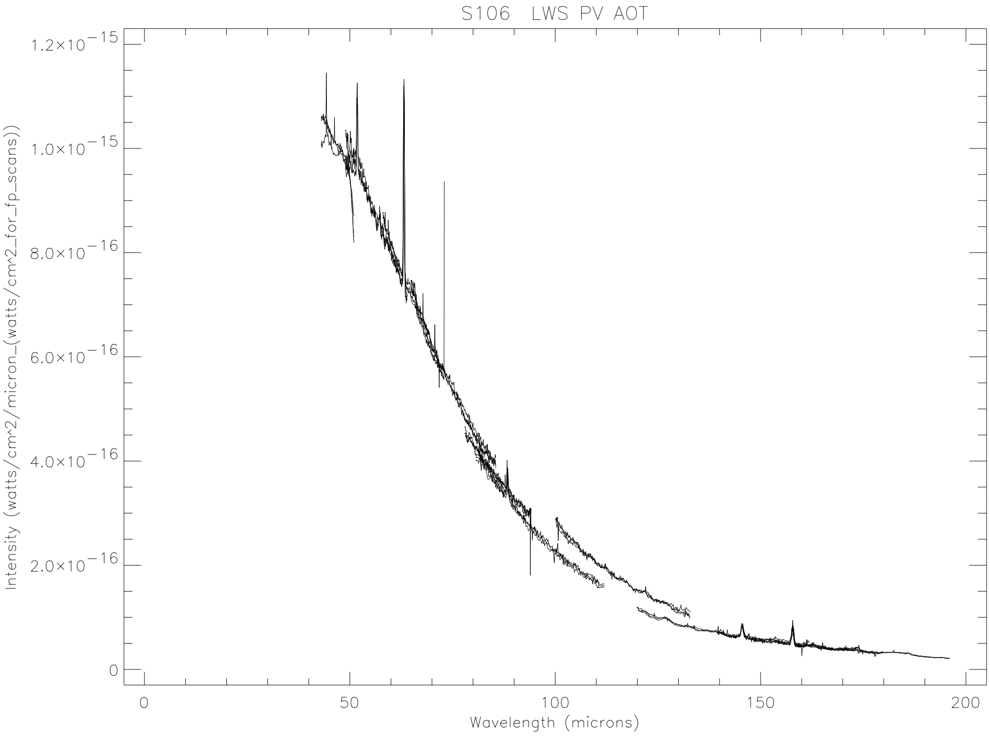

Figure 3. *The same observation as in figure 2 , but processed with a pipeline version from routine phase. The different subspectra fit here much better together.*

signals as a function of time is used to correct for the responsivity drifts within wavelength range AOT's.[6] For weak sources it can make a big difference whether the dark signal is subtracted before or after this drift correction, since in the first case the algorithm can produce significant distortions of the signal.

4. PROCESSING BEYOND AUTO ANALYSIS

The LWS Interactive Analysis Software[7] allows the observer to examine and to modify the products of the pipeline. It contains the following inspection routines:

- Display of the signal from the illuminator flash sequences and the measured dark current in astronomical observations: This enables the observer to identify possible corruption in the data (see section 2.3)
- Display of the absolute responsivity correction factors
- Display of pointing directions (this is particularly useful for raster observations)
- Comparison of the SPD results before and after pipeline deglitching
- Comparison of the SPD results with and without responsivity drift correction
- Comparison of SPD, Relative Spectral Response Function and AAR: This comparison is useful for the identification of spurious features in a grating spectrum.

With them it is possible for the observer to check that the crucial steps of the pipeline processing have been executed successfully. If he is not content with the execution of a particular processing step in the pipeline, he can use LIA also for repeating this step in an improved way, e. g. by using nominal values for the dark currents instead of the measured ones. It is possible, too, to create a product without the responsivity drift correction.

In order to facilitate the scientific analysis of the data another set of interactive routines was written: the ISO Spectroscopy Analysis Package.[8] This software uses AAR as starting point and performs the following tasks in Interactive Data Language (IDL):

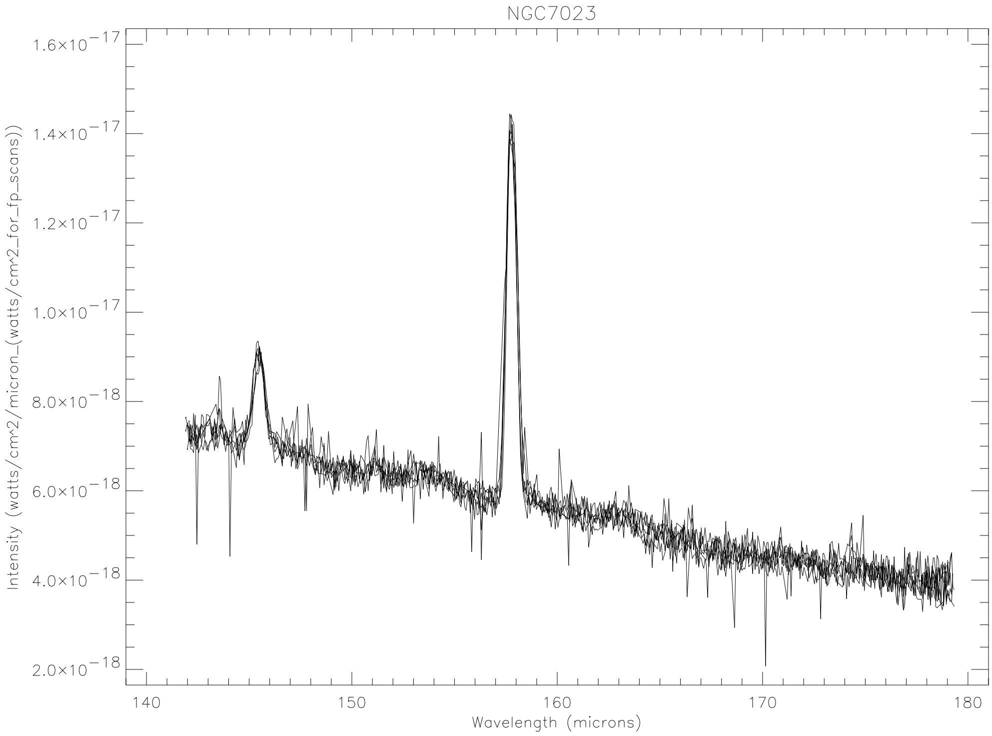

Figure 4. *Spectrum in one detector (LW4) of the reflection nebula NGC7023 as observed in revolution 448 and processed to Auto Analysis level with a pipeline version used during routine phase (v. 6). The [OI] line at 146 micron and the [CII] line at 158 micron are easily recognizable in the spectrum.*

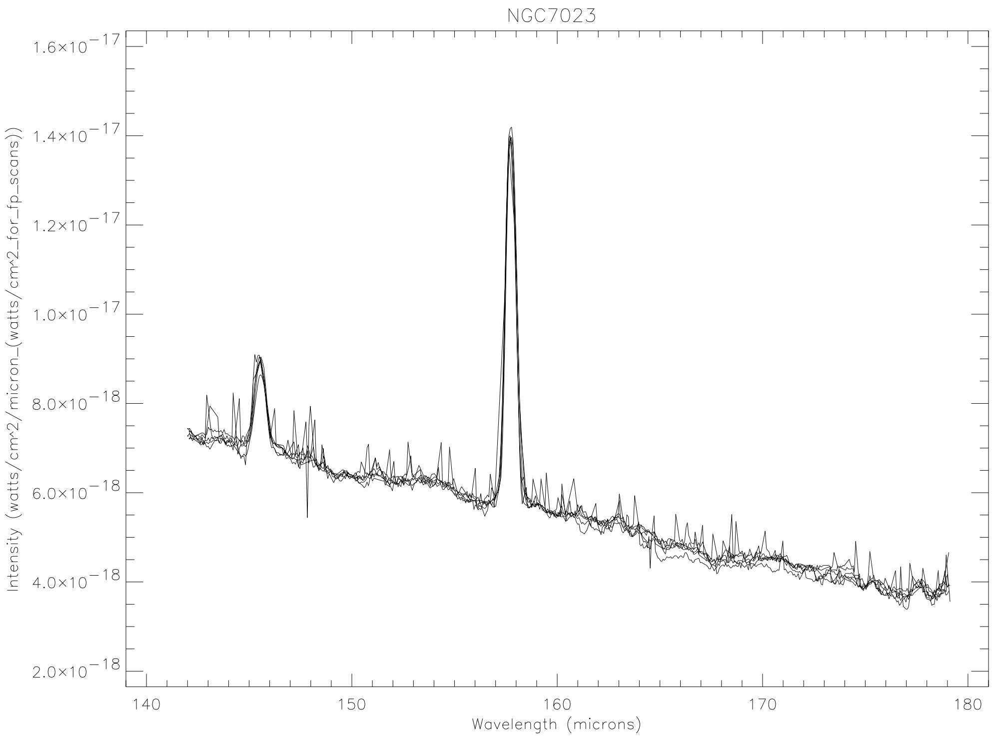

Figure 5. *The same observation as in figure 4, but processed with a prototype of the next pipeline version (v. 7). The spread of the measured values has been significantly reduced, mainly because of an improved ramp fitting.*

- Display of spectra in various units with the possibility to remove outliers
- Mathematical operations on the data like averaging of scans or subtracting of background observations
- Removal of fringes in the spectra that are caused by emission off the center of the field of view
- Removal of discrepancies between adjacent detectors by comparing their signals in the overlap region

5. LIMITS OF PIPELINE PROCESSING

Although the pipeline produces reliable results for the great majority of observations, there are some conditions which require special attention. The most important case is the AOT for Fabry Perot range scans (L03), where the data products should only be used for identification of lines. An insufficient accuracy in setting the grating to the correct position results in an uncertain contribution of the underlying grating profile to the FP spectrum. Therefore the flux calibration in the pipeline can produce here significant errors, and special interactive processing is required to obtain the correct shape and flux in the spectrum.

Other peculiarities of the instrument, for which there are currently no corrections in the pipeline, are:

- Memory effects in LWS data: The transient response of the doped-Germanium photoconductors produces a shift of the line center and in certain cases also variations of the continuum shape depending on the scan direction.[9] The pipeline uses at present the average relative spectral response function of forward and backward scans, but separate RSRF's for each scan direction will be available in the future.
- Special flux calibration for strong sources: Pipeline processing assumes a linear relationship between photocurrents and flux, but this is not true for sources with fluxes of several thousand Jy or more. A strong source correction is being tested at the time of writing and may be available in a future version of the pipeline.[6]
- Fringes in the spectra: Reasonable defringing routines are available for post-pipeline processing.
- Flux calibration for extended sources: Since the beam size of the LWS differs from the model value and changes as a function of wavelength and direction in the focal plane, the accuracy of the flux calibration is reduced for extended sources. In order to quantify this problem, the signal from a point source at different positions in the field of view has been measured. But even when the beam geometry is exactly known, highest accuracy in the flux of an extended source can only be achieved interactively with certain assumptions on the size and the shape of the source.
- Spurious features in spectra due to filter leaks: This is, like the previous point, an instrumental effect that as a matter of principle cannot be resolved automatically. The observer has to remove the spurious features by fitting a Gaussian profile to them.

6. CONCLUSIONS

The experience gained with the LWS during flight made some changes to the pipeline processing of the data necessary. These modifications reflected the growing accuracy in the wavelength and flux calibration of the instrument and the improvements of its observing strategy. Each increase of calibration accuracy can be implemented by a simple update of a table; changes to the processing itself are done by modifications of subroutines. This structure has proven to be flexible enough to adapt quickly to the altered behaviour of this complex instrument in space, and it was possible to obtain with automatic processing already during the mission scientifically valid data for the great majority of observations. Refinements of the calibration in the post mission phase will result in further updates of the pipeline during the next few years.

ACKNOWLEDGMENTS

ISO is an ESA project with instruments funded by ESA Member States (especially the PI countries: France, Germany the Netherlands and the United Kingdom) and with the participation of ISAS and NASA. The LWS was designed and built by a consortium of scientists and engineers from Canada, France, Italy, the UK and the USA. For the software systems discussed here the majority of work was performed at ESTEC SAI and the Rutherford Appleton Laboratory (RAL).

REFERENCES

1. P. E. Clegg, et al., "The ISO Long-Wavelength Spectrometer," *A&A* **315**, pp. L38-L42, 1996.
2. M. F. Kessler, et al., "The Infrared Space Observatory (ISO) mission," *A&A* **315**, pp. L27-L31, 1996.
3. K. J. King, et al., "The life enhancement possibilities of trend analysis," these proceedings, 1998.
4. B. M. Swinyard, et al., "In orbit performance of the ISO long wavelength spectrometer," these proceedings, 1998.
5. M. J. Burgdorf, et al. , "In-Orbit Performance of the LWS Detectors," in *The Far InfraRed and Submillimetre Universe* , **ESA SP-401**, ed. A. Wilson, pp. 353-356, 1997.
6. B. M. Swinyard, et al. , "Calibration of the ISO Long Wavelength Spectrometer: Correction of Detector Responsivity Drift and Non-Linearity," in *The Far InfraRed and Submillimetre Universe* , **ESA SP-401**, ed. A. Wilson, pp. 441-444, 1997.
7. S. D. Sidher, B. M. Swinyard, A. S. Harwood, S. D. Lord, and S. J. Unger , "The LWS Interactive Analysis Software," in *Proceedings of the First ISO Workshop on Analytical Spectroscopy* , **ESA SP-419**, ed. A. M. Heras, K. Leech, N. R. Trams, and M. Perry, pp. 297-298, 1998.
8. E. Sturm, et al., "The ISO Spectral Analysis Package ISAP," *ASP Conf. Ser.*, 1997.
9. T. Lim, B. M. Swinyard, X.-W. Liu, M. Burgdorf, C. Gry, S. Pezzuto, and E. Tommasi, "Memory Effects and the LWS RSRF," in *Proceedings of the First ISO Workshop on Analytical Spectroscopy* , **ESA SP-419**, ed. A. M. Heras, K. Leech, N. R. Trams, and M. Perry, pp. 281-282, 1998.