Chapter 13

Assessing NEWSIPS Data Quality

The processing techniques described in this manual provide some information on the quality of the image and the extracted spectral data which was not available with the previous IUESIPS processing. This information can be found in the processing HISTORY portion of the FITS label for each image. Each of these parameters is described below. In some cases, these output parameters can alert the user to the fact that the raw image data are corrupted or that the distortion of a particular image with respect to the ITF is unusually severe. Such problems are inherent in the raw data and cannot be overcome with processing techniques. Other parameters alert the user to difficulties with the extraction of the spectral data. In these cases, alternate extraction methods customized by the user may yield better results.

The "ITF cross-correlation parameters" (Item 3) refer to statistical information based on data obtained from the entire photometrically corrected region of the image. While these numbers can be useful for quick-look evaluations, the spatial data represented must be examined in two dimensions to understand fully the distortion characteristics of a particular image. These warnings therefore should be used with caution for anything but a gross categorization of image quality.

1. Raw Image Data Number (DN) levels:

   These estimates are determined automatically during the raw-image screening process using the algorithms described in Chapter 4.8. These values may differ significantly from the estimates given in the RA comments field in the database, which is based on a manual estimate at the time the exposure was read down. The estimate of the exposure levels in the RA comments field is made in slightly different ways at GSFC and at VILSPA and is not necessarily consistent from image to image. The NEWSIPS background and continuum DN levels, on the other hand, are consistent for the entire archive.

   Maximum continuum. This parameter gives an estimate of the exposure level of the raw spectral data. The maximum continuum DN level can be used to determine relative exposure differences between various observations of the same object. Of course, the spectral morphology of the object must be taken into account in evaluating the ex-

posure level of an image. In particular, one must be careful when using these numbers if they were determined from emission line objects. Contamination of the continuum level reading from an emission feature may occur in some instances when there is a large spectral format shift (e.g., due to target centering errors or objects with large redshifts).

Mean Background. The mean background DN level for a short exposure taken during a time of low radiation is about 20–40 DN. This is then the minimum background one would expect. Long exposures and exposures taken on high radiation shifts will have higher mean background DN levels. If the maximum continuum DN level is not at least 50 DN above the mean background DN level, the spectrum will most likely contain little useful data unless it is an emission line source.

2. High-dispersion order registration warnings:

   Several warning conditions can occur during the high-dispersion order registration process which alert the user to a potential error in the spatial alignment of the science image relative to the fiducial. In each case, this may lead to a loss of flux in the extracted spectrum.

   Insufficient flux for empirical order registration. RMS of found vs. actual order positions > noise criterion. Predictions based on time and temperature will be used. Indicates that the average RMS of the differences between the found and fiducial order positions exceeds 1.5 pixels or is equal to zero. The former condition usually occurs as a result of a lack of flux in the spectral orders, while an RMS of zero can be produced by images with heavy saturation or unusually large numbers of missing minor frames (MMFs). A potential error in the alignment of the high-dispersion resampled image (SI) with respect to the fiducial image may take place, as use of the predicted time and temperature motions do not take into account target centering errors or the use of an offset reference point when placing the target in the aperture. As a result of the misalignment of the high-dispersion SI, a loss of flux may be seen in the extracted spectrum.

   Insufficient flux for empirical order registration. Too few valid orders found. Predictions based on time and temperature will be used. Denotes that the total number of orders found is less than 3 (5 for the SWP). The potential errors involved in the use of predicted time and temperature motions are the same as those described in the previous condition.

   Global offset nn pixels relative to fiducial. Relative order locations defaulted to those of the fiducial. Signals the user that the distribution of the orders is incorrect and that only the global offset will be used. Usage of the global offset for alignment of the high-dispersion SI does not take into account differential expansions and contractions of the order separations. As a result, minor miscenterings of the extraction slit for orders at the short- and long-wavelength ends of the camera may occur resulting in a slight loss of flux for particular orders.

Found offset exceeds threshold of 4 pixels. Predictions based on global offset will be used. This condition happens when the maximum shift for an individual order exceeded the threshold value. Use of the global offset may yield extraction errors as defined in the previous condition. This error condition is only applicable to the long-wavelength cameras.

3. ITF cross-correlation parameters:

   These numbers are generated during the cross-correlation portion of NEWSIPS processing and are usually a good indicator of the signal-to-noise ratio (S/N).

   Percent successful cross correlations. The percent of successful cross correlations for each image is recorded in the HISTORY portion of the FITS header. For most low-dispersion images, greater than 95% of the attempted cross correlations are successful, according to the criteria described in Chapter 5. If less than 90% of the cross correlations were successful, the image either suffers from unusually large local or global distortions, the raw image background is heavily saturated, the image is a partial read, or the raw image data are corrupted. For high-dispersion data, the above stated thresholds are much lower due to the smaller area background available for cross correlations. The majority (∼75%) of high-dispersion images have 80% or better successful cross correlations.

   Median cross-correlation coefficient. The median cross-correlation coefficient for an image is the median of the cross-correlation coefficients for all patches of the image. In general, the best signal-to-noise in a low-dispersion extracted spectrum can be expected when this parameter is greater than 0.7. When this parameter is less than 0.6, the S/N of the photometrically corrected data, and ultimately the low-dispersion extracted spectrum, may be degraded. In high dispersion, this number tends to be lower for optimally exposed images; most high-dispersion images have a median cross-correlation coefficient of 0.4 or better.

   Mean shift. This parameter is the mean of the magnitudes of the shift vectors for all patches used in the cross correlation. In general, the mean shift is inversely correlated with the median cross-correlation coefficient. Mean shifts of less than 0.5 pixel tend to yield the best S/N. This value holds true for both high- and low-dispersion images.

   Maximum shift. The maximum shift is the largest magnitude of a shift vector in the array of patches used for the cross correlation. This value can be misleading as it may represent a correlation with a correspondingly large displacement which, during the filtering portion of the processing, is declared "invalid" (i.e., it does not conform to the overall trend of spatial deviations in the data). A low-dispersion image having a maximum shift of more than 1 pixel may suffer from unusual local distortions. Such local distortions can signify regions of the spectrum which may have degraded S/N. On average, high-dispersion images have maximum shifts of approximately 2 pixels.

4. Camera Temperature:

   It is known that images acquired at camera head amplifier temperature (THDA) readings of more than a few degrees from the mean THDA of the ITF images almost always suffer from unusually large distortions compared to the majority of the images in the archive. The cross-correlation algorithm has a significantly higher confidence in the pattern recognition and the mean shifts are smaller when the THDA of the image is close to that of the ITF (see Table 13.1). Images with THDAs more than a few degrees from the mean ITF THDA for that camera will probably have a poor S/N. Note that the THDA is a secondary indicator of image registration quality; the median correlation coefficient is a more direct measure of registration quality.

Table 13.1: Mean ITF THDAs

Camera	LWP	LWR	SWP
ITF	1992	1983	1985
THDA	9.6	14.5	9.3

5. High-dispersion background determination warnings:

   During the extraction of the background fluxes, various warning/failure messages may be produced, which point towards errors in the calculation of the background fluxes.

   Background determination failed due to insufficient flux; background flux set to zero. This failure most likely indicates that a large portion of the image is affected by MMFs. Therefore, the background extraction algorithm has determined that background fluxes are not reliably sampled.

   Flare detected. This condition only occurs for the LWR camera and signifies that the background determination module found a flare. The presence of a flare will most certainly lead to localized errors in the background solutions. The potential also exists for the effect of the flare to propagate across much of the image.

6. High-dispersion boxcar extraction warnings:

   Several warning messages may be generated during the extraction of high-dispersion fluxes and are indicative of abnormalities in the centroiding of the extraction slit.

   Order nnn found at > 0.5 pixel position from line xxx.xx - yellow light. nnn is 100 for the SWP and 90 for the LWP and LWR. *xxx.xx* is 290.74 for the SWP, 412.71 for the LWP, and 404.20 for the LWR. This condition is triggered when the appropriate "checkpoint" order (either 90 or 100 depending on the camera) deviates half a pixel or more from its fiducial (expected) location. These "checkpoint" orders are selected to be representative of the image as a whole, and fall in regions of the image where the camera

sensitivity is reasonably high. An occurrence of this warning alone is not necessarily indicative of a serious problem; however, the user should be aware that the apparent location of this "checkpoint" order is beyond the normally expected bounds and may indicate a larger than normal misalignment of the high-dispersion SI relative to the fiducial. The LWP processing HISTORY initially reported the line position for order 100 (sample position 316.02). This was subsequently changed to order 90 after the start of the processing effort, as this order is in a region of higher sensitivity than order 100 for the LWP and LWR. This change only affects LWP and LWR high-dispersion images processed after July 28,1997.

Order nnn explicit centroid determination invalid. Fiducial centroid used. Occurs when there is a lack of flux in a particular order. As a result, the positioning of the extraction slit defaults to a predetermined order position which may result in a loss of flux.

Order nnn spectrum centroid found beyond tolerance region. Fiducial centroid used. Indicates that the center of the order was found outside a preset limit which varies from 3.0 pixels to 0.5 pixels depending upon the order number. This warning is indicative of either a lack of flux in this order or a potential misalignment of the high-dispersion SI relative to the fiducial. As a result, the positioning of the extraction slit defaults to a predetermined order position which may result in a loss of flux.

7. Low-dispersion *SWET* extraction parameters and warnings:

   The following messages appear in the HISTORY position of the FITS header and alert the user to potential problems with the flux extraction.

   Spectrum centroid displacement from predicted center. The centroid of the spectrum was found more than 2 pixels from the predicted center of the spectrum, based on time and temperature dependencies. In this case, it is likely that the image suffers from unusually large spectral format shifts or that there were large target centering errors. As a result, the wavelength calibration could be in error due to a global displacement along the dispersion or local distortions not corrected for in the wavelength linearization and wavelength calibration.

   Spectrum peak displacement from the spectrum centroid. The peak of the flux in the spectrum was found more than 1 pixel from the centroid of the spectrum, in the direction perpendicular to the dispersion. The spectrum may not be a point source and may benefit from re-extraction with a wider slit.

   Average peak FN. Images with an average peak of less than 5 FN revert to the use of a default extraction center and profile. These images suffer from very poor signal strength and should be used with caution.

   Size of cross-dispersion profile bins. A bin size of greater than 1 pixel may indicate that a large fraction of the spectral data are bad, missing, or have little or no flux above the background level. In most cases, these images revert to the use of a default extraction profile. Extracted data should be examined carefully.

Number of spline nodes. Spline nodes are used to determine the shape of the spectrum along the dispersion direction. Spectral data which require 2 or 3 spline nodes have a very low overall signal strength or are quite noisy, and may have significantly lower than average S/N. Spectral data utilizing only 2 spline nodes are, in fact, forced to revert to use of a default profile. Spectral profile fitting is discussed in Chapter 9.4.

Percent of flagged pixels in the extracted spectrum. If more than 10% of the extracted points in the MXLO file are flagged as either bad or cosmic ray hits, the error flags and the sigma vector should be examined carefully for the image.

8. Non-standard read and/or exposure gain:

   The photometric quality of images acquired using non-standard read/exposure gains is not well known. Careful examination of this type of data is recommended prior to its use. See Chapter 11.1.3 for a discussion of the gain correction factors.

9. Extrapolated wavelength calibration and/or time-dependent sensitivity degradation correction:

   The image was acquired at a time outside the range of the dates used to generate the wavelength calibration or sensitivity degradation correction. These calibrations were extrapolated in time and could in principle be slightly in error. Note, however, that analysis of late-epoch data by Garhart (1997) has shown that the extrapolations of the LWP point/extended/trailed and SWP trailed time-dependent sensitivity degradation corrections are nonetheless valid.

10. Extrapolated temperature-dependent sensitivity degradation correction:

    The THDA at which the image was acquired is outside the range of temperatures used to derive the correction. Flux corrections performed using extreme THDAs could be slightly in error.

11. Data quality (ν) flags:

    These flags indicate abnormal conditions in the data which can range from fairly minor to quite serious situations. The ν flags for the merged extracted image should be examined carefully in order to ascertain whether or not a particular data point is good or bad. In general, ν flag values of -8 or more negative (bits 4–15) are indicative of unreliable data. See Chapter 3 for a detailed discussion of ν flags.

    More error conditions are flagged in NEWSIPS data than were flagged in IUESIPS data. In addition, *all* error conditions that affect a pixel in the two-dimensional data are bit-encoded into the ν flag data, while only the most severe error condition affecting a pixel could be recovered from IUESIPS data. The attentive investigator will have a better understanding of the errors inherent in the NEWSIPS data than was possible with IUESIPS.

    It is particularly important to consider the ν flag values when analyzing extracted low-dispersion spectral data. The ν flag assigned to each extracted point in low dispersion

is determined from statistical considerations and does not represent *all* error conditions for the pixels used in the calculation of the extracted spectral point. Examination of the low-dispersion resampled ν flag image allows recovery of this error information and should always be performed in order to determine the locations and nature of the error conditions as identified on a pixel-by-pixel basis. See Chapter 9.6 for details concerning the generation of the one-dimensional (1-D) ν-flag spectrum in low dispersion. In high dispersion, *all* ν flags assigned to pixels used in the calculation of the extracted spectral point are carried over to the 1-D ν-flag spectrum.