The HST Treasury Program on Eta Carinae

Technical memo number 11 -- February 2006

RELEASE NOTES FOR VERSION 1.2.1 OF THE DATABASE

John C. Martin, Kris Davidson (UMN), K. Ishibashi(MIT), & Roberta M. Humphreys (UMN)

University of Minnesota

(If you use information or advice from this memo, please acknowledge it and the net site http://etacar.umn.edu in any resulting publications; thanks.)

1. Introduction

This is our first unrestricted public release of data gathered by the HST Treasury Project for Eta Carinae. At this time the publicly-available archive includes all of the "Version 1.2" HST/STIS/CCD spectroscopic data images, covering the approximate wavelength range 170—1000 nm at epochs from January 1998 to March 2004. STIS/MAMA UV echelle data will be added during summer 2005. STIS/ACS images and ground-based VLT/UVES spectroscopy will become available here somewhat later.

The HST/STIS CCD data include several improvements over the normal STScI pipeline and standard CALSTIS reductions. Those are outlined in Section 2. Section 3 lists the caveats and warnings for users of the database. And Sections 4 and 5 describe software tools developed for use with this data and the plans for future releases of the database.

This release does not represent the final data reduction. Additional processing and software will become available as time progresses. For more information on the full extent of the HST Eta Carinae Treasury Project please visit our web site (http://etacar.umn.edu).

When you use data from this archive in a publication, please mention it in the text and include the following acknowledgment or equivalent wording:

"This paper has used data from the Hubble Treasury Program for Eta Carinae, http://etacar.umn.edu, supported by grant GO-9973 from the Space Telescope Science Institute. The STScI is funded by NASA."

This document will be archived at http://etacar.umn.edu/treasury/techmemos/

2. Items of Interest

This document contains all the notes that pertain to the release of Version 1.2 including those already outlined in the Version 1.0 release. For historical reference, the Version 1.0 Release Notes can be found on our web site at:
http://etacar.umn.edu/treasury/techmemos/pdf/tmemo004.pdf
The following items pertain to the release of Version 1.2 of the database:

a) *File Naming Scheme*
The naming scheme used in Version 1.2 is fairly simple, and is described in detail in Technical Memo #3 on our web site:
http://etacar.umn.edu/treasury/publications/pdf/tmemo003.pdf
Each file name clearly indicates the observation date (e.g. 2003.51) and other information. Technical Memo #5 outlines the details surrounding the changes we have made in Version 1.2 including a table which cross references the names used in previous versions of the database with the current names:
http://etacar.umn.edu/treasury/publications/pdf/tmemo005.pdf

b) *Improved Pixel Interpolation*
Improved techniques used to model and interpolate the pixels in the STIS CCD are outlined in Technical Memo #1 on our web site:
http://etacar.umn.edu/treasury/publications/pdf/tmemo001.pdf
This procedure has improved the effective resolution of the STIS CCD on the dispersion and cross-dispersion axis since the previously used scheme effectively acted as a *position dependent* blurring function. Data reduced by the new method are of a quality which surpasses other available data processing techniques for STIS CCD data. The "scalloping" which is obvious in narrow extractions from data reduced by the former methods is almost completely eliminated by this new method. See the illustration in Appendix A and Technical Memo #1 for full details. Also note that the reduced data have been rebinned so that one pixel in the raw data corresponds to two pixels in the reduced data (the effective pixel scale in the reduced data is about 0.025035 arcsecond/pixel).

c) *Initial Bad/Hot Pixel Removal*
Bad pixels (hot pixels, cosmic ray hits, etc.) were identified in this data by two independent techniques. First, we applied the standard CR-SPLIT method which is familiar to HST users. The standard CR SPLIT method requires at least two exposures (a condition that was not always met) and can miss a few bad pixels that appear in all the individual frames. Therefore, we developed a second technique which identifies bad pixels without requiring multiple exposures. Our method relies on the fact that a legitimate point source is sampled by several pixels on the STIS CCD. Each pixel is compared with its neighbors in 5x5 square centered on that pixel. After rejecting the two highest and lowest neighboring values, we calculate the mean (V_{avg}) and the r.m.s. dispersion (D) of the 20 remaining comparison pixels. The R value of the pixel being tested (V_0) is assessed by the expression: $R = \left|V_0 - V_{avg}\right| / D$. In the cases where R exceeds a

well-chosen threshold (R_c) the pixel is flagged as bad. By iterating this process, sizeable clusters of bad pixels can be identified and eliminated. This method was carefully tuned and tested so that it eliminates nearly all the obviously bad pixels in a single exposure without adversely altering the noise structure of the data or the data itself.
A more detailed technical memo on this technique will be made available at a later date.

d) *Flagging Bad Pixels Which Result from Exposure Saturation*
Nonlinear interactions between the distortion correction and the pixel interpolation technique make it a non-trivial task to identify pixels in the final reduced data which may have been adversely affected by pixels that are saturated by overexposure in the raw data. Unfortunately, one cannot rely on the Data Quality (DQ) array in the FITS files to properly identify all of the bad pixels (see 3a in Caveat's & Warnings section below). A list of the observations which have over-exposed pixels (identified from the raw data files) is available on our web site at:
http://etacar.umn.edu/~martin/etacar/sattab.html
In many cases the raw exposure with the saturated pixels has a shorter companion exposure with the same grating tilt and target at the same epoch. In those cases, we were able to compare the reduced data from the long exposure (with the saturated pixels) to the shorter exposure (without saturated pixels) in order to determine which pixels have been adversely affected in the longer exposure. Very conservative tolerances were used to identify the bad pixels in the longer exposure. The pixels identified as having been adversely affected by overexposure have been assigned a value of NaN in the flux data array so that it is unavoidably obvious that those pixels are bad.

> *We refer to NaN as defined in IEEE 754. We use a quiet NaN (QNaN) which is explicitly generated by the expression: NaN=(-999./0.0).*

Twenty of the overexposed observations have no companion short exposure. As a result, we cannot easily identify the pixels which are adversely affected by overexposure in those observations. A history entry has been added to the primary and FLUX HDU of these FITS files to note that they possibly contain bad pixels which have not been properly identified.
Technical Memo #8 recounts the full details concerning this procedure:
http://etacar.umn.edu/treasury/publications/pdf/tmemo008.pdf

e) *Patching of Saturated Hα and Hβ Exposures*
The STIS CCD observations of the Hydrogen Balmer lines Hα and Hβ are pairs of one long and one short exposure which intentionally overexpose the core of the Balmer lines from the central star in the long exposure in order to obtain greater signal in the wings of the lines and the fainter parts of the surrounding nebula. We have combined these exposure pairs into single images with a larger combined dynamic range than a single STIS CCD exposure allows. Each of these higher level science products contains an extra FITS image header (HDU #5 [MASK]) which is the array of weights used to combine the pixels from the long and the short image. In the case of these combined images the error (ERR) array contains the standard deviations of the weighted individual pixel values that went into building the flux array. See Technical Memo 8 for details:
http://etacar.umn.edu/treasury/publications/pdf/tmemo008.pdf

f) *Wavelength Calibration*

The wavelength calibration has been rigorously verified by measuring the wavelengths of narrow stellar photospheric spectral features in standard stars reduced in the same manner as the Eta Carinae data. The current wavelength calibration meets or exceeds the standard set by the StScI pipeline and CALSTIS. Remaining wavelength errors may result from imperfect pointing of the STIS slit.

A number of the datasets have no WAVECAL associated. Instead the wavelength scale was determined by cross correlating with a known feature (i.e. lines in the Weigelt knots). See section 5a.

g) *Removal of the WAVE Arrays*

The WAVE array HDU in the FITS files has been removed from the files in this database. We have done this because: 1. the WAVE HDU contained redundant information also stored in the WCS, 2. the WCS information is a more space efficient and precise means of saving the wavelength calibration, and 3. we want to avoid any possible confusion between the WCS information and the lower precision WAVE HDU.

We are aware that some existing analysis software may rely on having a WAVE array HDU in the FITS file. Therefore, we are providing software that will rebuild the WAVE array using the WCS information. Both an IDL subroutine and C source-code for this procedure can be obtained online at:

http://etacar.umn.edu/treasury/software/addwave

h) *Flux Calibration*

The flux calibration of the data was checked by comparing standard stars BD +75 325 and AGK +81 266, reduced by our technique, with data published by Bohlin et al. (2001). The current flux calibration meets or exceeds the standard set by the StScI pipeline and CALSTIS for extended source spectral flux calibration across the entire range of wavelengths and epochs covered by this data.

A possible point of confusion regarding the flux array has been clarified by adding notes to the FITS headers. The units in the FLUX array (as given by the BUNITS keyword) are ergs/sec/cm^2/Å/reduced-row. "Reduced-row" refers to a single row in the reduced data array **NOT** an original CCD row. We have chosen this convention so that the flux can be integrated across the rows in the FLUX array without applying a scalar correction to account for the pixel interpolation.

However, the units on the ERR array are different from the FLUX array. They are in ergs/sec/cm^2/Å/original-CCD-row. See section 3d for further discussion concerning the status of the FITS ERR arrays.

It should be noted that when a 1D spectral extraction is made from the 2D spectra in the database that extracted spectrum is still uncorrected for the extraction height. These corrections will be made available as software in the near future. See Section 5b for details.

i) *Centering of the Nominal Target on the Cross-Dispersion Axis*

We have determined the location (fractional row number) of the nominal slit center for each observation. The following algorithm was used to determine the reference row for each observation:

1) If the central star is present on the slit, then its actual position is used as a reference point to determine the row for the slit center.
2) If the slit for an observation does not include the central star but that observation is immediately preceded or followed by a matching observation with the same grating parameters which has its reference row determined by method #1, then the reference row from the matching observation is used.
3) If neither of the first two methods apply, then a plot of the cross dispersion axis the observation is matched with a slice extracted from HST ACS/HRC images to determine the reference rows.

The WCS information in the FITS headers has been edited so that the CCD row where the nominal target appears is identified by the keyword CRPIX2. Also, as per the findings of Bowers & Baum (1998), the cross-dispersion pixel scale has been set to 0.025035 arcseconds/reduced-pixel (CD2_2 FITS header keyword). As a result, the WCS information should return the proper coordinate along the slit, with the nominal target located at zero. In most cases, the centering of the nominal target is accurate to better than ±0.0127 arcseconds (half a reduced pixel).

It is important to note that some observations which are nominally targeted on the Weigelt blobs are not centered on those features. That is not to say that the slit is not oriented to include the nominal target. However, the slit may not be centered on that target, meaning that the reference row does not pass through the nominal target.

3. Caveats and Warnings

a) *In Some Cases There is no WAVECAL*

There is no formal WAVECAL (STIS/CCD wavelength calibration) associated with 92 of the observations in the data archive. In those cases the zero point for the wavelength calibration is uncertain. The cases where this occurs have been noted by adding comments to the CRVAL1 and WOFFSET values in the FLUX HDU of the FITS files which clearly state that there is "NO WAVECAL" for that dataset. They are also flagged in the database so that a user can tell if a dataset has no WAVECAL.

Ninety two (92) of the STIS CCD spectra were taken with no WAVECAL calibration data sets in order to optimize the scientific return in a limited number of available HST orbits. In most cases we have been able to calibrate the affected data using narrow, stationary nebular emission features (e.g., Weigelt Blobs). Note that the stellar spectrum (or the reflection of it) is not a suitable source for wavelength calibration; hence we avoid using it for the purpose.

In 78 of the 92 cases we used the narrow emission spectra of the Weigelt B and D blobs to calibrate the wavelength scale of the STIS CCD. In those cases the Weigelt blob spectrum was calibrated using the data 1998 March 19 where they were observed with roughly the same slit orientation. This class of calibration includes observations made on: 1999 February 21 (files beginning with c914), 2000 March 20 (files beginning with cA22 numbered 0010 through 0350) and 2004 March 7 (all files beginning with cE18).

For the rest, there are no WAVECAL or Weigelt knot spectra. Therefore we have used fainter nebular emission lines (sometimes the reflection nebula) to determine the zero wavelengths. This type of calibration can be very uncertain and should not be implicitly trusted under any circumstances. This affects the following datasets: 1998 November 25 (files beginning with c890 numbered 0010 to 0110), the G230MB (cenwave= 1713AA) spectrum in the visit on 2000 March 13 (file cA20_0010), and STIS CCD data in the visit on 2000 October 9 (files beginning with cA77).

b) *Do Not Entirely Trust the Data Quality (DQ) Array*

At present, the data quality (DQ) array in the FITS files follows a convention established by the STIS Instrument Development Team (implemented in the latest versions of CALSTIS) rather than the convention established in the STIS Instrument Handbook. This means that only the "worst" thing that occurred to a pixel during reductions is recorded in the DQ array rather than an audit trail of all of the errors encountered by that pixel. We also have some justifiable concerns that the DQ array might not properly identify all the pixels in the reduced image which have been adversely affected by overexposed pixels in the raw image. As a result, we suggest that users consider the DQ array a best guess as to the status of a pixel and place more trust in flagging of bad or overexposed pixels which has been done by us post-production.

We hope to improve on this situation in a later release.

c) *Beware Possible PSF Changes on the STIS CCD*
The PSF of a point source on the STIS CCD is variable and dependent on grating, spectral wavelength, and column number on the CCD. We plan to address these issues in a later release of the database. In the mean time, we have documented several of these effects on our web site in Technical Memo number 2 (http://etacar.umn.edu/treasury/publications/pdf/tmemo002.pdf).

d) *The Error Array Does Not Account For All Possible Errors and May Be Unreliable*
The values in the error (ERR) array are the theoretical statistical errors for each corresponding pixel calculated by the method outlined on page 12 of the CALSTIS Users Manual. We have verified that these values are roughly equivalent to the average noise computed from the pixel values in the unexposed portions of the CCD. However, users should be aware of the origin of these values and resist the temptation to over-interpret them since when pixels are rebinned by interpolation; the appropriate error value depends on the application. A full explanation of the contents of the ERR array is given in Technical Memo #7:
http://etacar.umn.edu/treasury/techmemos/pdf/tmemo007.pdf
Additionally, we have some concerns about the contents of the ERR array in the rectified two-dimensional spectra produced by the standard pipeline at the STScI online MAST archive. There appears to be some inexplicable "pattern noise" of unknown origin in those arrays (see Figure 1 in Technical Memo #7). This coherent noise is not present in the ERR arrays we have produced.

e) *Scattered Light/Diffuse Background and Ghost Images Are Still Present*
The scattered light/diffuse background has not been removed from the data. The level of scattered light rises with increasing wavelength and can be significant at long wavelengths in exposures which are long and/or have a very bright source on the slit. We are developing software to remove the scattered light from a bright point source (see Section 5d).
There are also ghost images in the STIS CCD optical system (Gull et al., 2002) which are still present in the data. The most obvious place this occurs is in exposures of the central star of the Hα line. In these cases the ghost appears below and to the right of the actual feature (decreasing row number, increasing column number). Do not mistake this ghost for diffuse emission from the surrounding nebula! We have produced software which removes the Hα ghost (see Section 4a).
This ghost is not to be confused with various anomalies measured in the cross-dispersion profile (see Technical Memo #2 and Section 5a in this document). We intend to eventually provide software to remove this effect as well.

f) *Extracting the Spectral Data Using IRAF*
Technical Memo #9 details how to use common IRAF procedures to extract one-dimensional spectra from the two-dimensional Treasury Project FITS data:
http://etacar.umn.edu/treasury/techmemos/pdf/tmemo009.pdf

4. Software Tools

There are a number of software tools which we have developed for use with the archive data. Furthermore, these tools might also be applicable for use with other HST/STIS data. Each of these software packages can be downloaded from:
http://etacar.umn.edu/treasury/software/

a) ghostbgone: *Removal of the Hα Ghost Image*
We have developed an empirical method to remove the ghost image of the bright Hα feature from the central star. See Technical Memo number 10:
http://etacar.umn.edu/treasury/techmemos/pdf/tmemo010.pdf

b) exfits: *Performing 1D Spectral Extractions*
We have written a software package, called *exfits*, to make 1D spectral extractions from the 2D spectra in the database. This software is available for download from:
http://etacar.umn.edu/treasury/software/exfits

c) addwave: *Reconstructing the WAVE array from WCS information*
We have removed the WAVE array HDU from the FITS files in our archive (see Section 2g). This software will rebuild the WAVE array using the WCS information. This software is available as an IDL subroutine or C source-code online at:
http://etacar.umn.edu/treasury/software/addwave

d) scatrm: *Background/Scattered Light Removal*
We have developed a rigorous well tested model for the background/scattered light in the STIS/CCD optical system from actual data (not first principles). The software to apply this model to the STIS/CCD data will be available in the near future. See the draft version of Technical Memo #6 online for details:
http://etacar.umn.edu/treasury/techmemos/pdf/tmemo006.pdf

5. Plans for Future Releases

We plan future releases of the database which will include the following items currently being developed:

a) *Deconvolution of the Cross-Dispersion Point Spread Function (XSF)*
Technical Memo number 2 (http://etacar.umn.edu/treasury/publications/pdf/tmemo002.pdf) details the behavior of the XSF on the STIS CCD. In principle we can deconvolute this function from the data in order to obtain even better spatial resolution on the cross-dispersion axis.

b) *Extraction Height Correction*
When a 1-dimenstional spectral extraction is made from 2-dimenstonal data in the archive, the resulting spectrum lacks a correction for the extraction height (number of rows included in the extraction). We are currently calibrating this correction and when we have finished it will be available as a piece of software.

c) *Spliced Spectra*
We plan to combine the different grating tilts into a single spliced spectrum that spans the entire wavelength range observed at each epoch in the archive. This task is not very difficult for 1-dimensional extracted spectra, but the 2-dimensional case is still largely unexplored.

d) *HST/STIS MAMA data and VLT/UVES data*
We plan to eventually include new reductions of the HST/STIS MAMA data which cover the same epochs as the HST/STIS CCD data. We will also eventually be adding ground-based data gathered with the VLT/UVES as part of an on-going observing campaign.

6. References

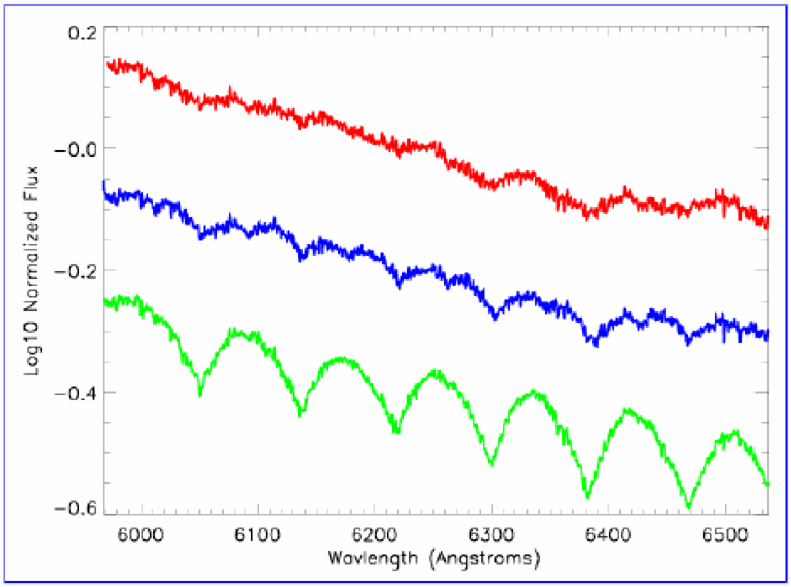
Bohlin, R.C., Dickinson, M.E., & Calzetti, D. 2001, AJ, 122, 2118.

Bowers, C.W. & Baum, S. 1998, "Plate Scales, Anamorphic Magnification & Dispersion: CCD Modes," STIS Instrument Science Report 98-23, (StScI: Baltimore).

Gull, T., Lindler, D., Tennant, D., Bowers, C., Grady, C., Hill, R.S., and Malumuth, E., 2002, "The STIS CCD Spectroscopic Line Spread Functions" presented at the 2002 HST Calibrations Workshop. (S Arribas, A Koekemoer, and B. Witmore, eds.)

Appendix A

The following demonstrates the effectiveness of the improved pixel interpolation method to avoid "scalloping" in narrow extractions on or near a bright source.



The above plot is of three extractions from a spectrum of BD +75 325 which are 0.1" wide centered 0.1" off the peak of the PSF. The bottom extraction was made from data reduced by the STScI pipeline (green). The middle extraction was made from data reduced by the STIS IDT software (blue). The top extraction was made from data reduced by our technique. This section of the spectrum of BD +75 325 is featureless so variations in each extraction are caused by "scalloping" introduced by the pixel interpolation methods. This illustrates that the new method we have employed is better than that used by the STIS IDT software by a factor of two or three (note the obvious periodic scallops in the middle/blue curve above) and constitutes a vast improvement over the standard pipeline processing.