Astronomy & Astrophysics manuscript no. ms
September 16, 2008
© ESO 2008

The Hubble Legacy Archive NICMOS Grism Data

Wolfram Freudling, Martin Kümmel, Jonas Haase, Richard Hook, Harald Kuntschner, Marco Lombardi, Alberto Micol, Felix Stoehr, and Jeremy Walsh

Space Telescope – European Coordinating Facility

submitted May 29, 2008; accepted Sep 3, 2008

ABSTRACT

The Hubble Legacy Archive (HLA) aims to create calibrated science data from the Hubble Space Telescope archive and make them accessible via user-friendly and Virtual Observatory (VO) compatible interfaces. It is a collaboration between the Space Telescope Science Institute (STScI), the Canadian Astronomy Data Centre (CADC) and the Space Telescope - European Coordinating Facility (ST-ECF). Data produced by the Hubble Space Telescope (HST) instruments with slitless spectroscopy modes are among the most difficult to extract and exploit. As part of the HLA project, the ST-ECF aims to provide calibrated spectra for objects observed with these HST slitless modes. In this paper, we present the HLA NICMOS G141 grism spectra. We describe in detail the calibration, data reduction and spectrum extraction methods used to produce the extracted spectra. The quality of the extracted spectra and associated direct images is demonstrated through comparison with near-IR imaging catalogues and existing near-IR spectroscopy. The output data products and their associated metadata are publicly available (`http://hla.stecf.org/`) through a web form, as well as a VO-compatible interface that enables flexible querying of the archive of the 2470 NICMOS G141 spectra. In total, spectra of 1923 unique targets are included.

Key words. NICMOS – infrared spectroscopy – Hubble Space Telescope, Calibration

arXiv:0809.2465v1 [astro-ph] 15 Sep 2008

1. Introduction

Three of the current Hubble Space Telescope instruments include built-in slitless spectroscopic modes: the Space Telescope Imaging Spectrograph (STIS), the Near Infrared Camera and Multi-Object Spectrometer (NICMOS), and the Advanced Camera for Surveys (ACS). The main advantage of slitless spectroscopy is that spectra can in principle be obtained from all objects within the field of view of an instrument. The main disadvantages are that the size of the object limits the achievable spectral resolution, that spectra might overlap, and that the background is relatively high because no slit mask prevents the full sky background from illuminating every pixel of the detector. In addition, extracting one-dimensional spectra from such data is a complex process and is often achieved using highly interactive procedures.

A fraction of the slitless data in the HST archive has been collected to obtain spectra of specific objects, but spectra of objects other than the primary targets have in most cases not been extracted or analysed. The goal of the ST-ECF HLA project is to extract spectra of all objects that have been observed with HST slitless spectroscopy modes and to serve these spectra through an archive with associated descriptions of the spectra, such as how much one spectrum is contaminated by those of nearby objects.

Each of the HST slitless spectrograph modes provides a different capability to HST. STIS and the ACS prisms cover the UV part of the spectrum that is not accessible from the ground. The NICMOS grisms cover IR wavelengths that can partially be observed from the ground, but where the background is much lower in space. The ACS optical channels covering the optical wavelength range also benefit from a combination of lower background and higher spatial resolution from space. The STIS grating provides the highest resolving power among the HST instruments with slitless spectroscopy modes.

The data analysis for each of the spectrographs presents a significant challenge, and specialised, mostly interactive tools are available for individual instruments such as NICMOS (Freudling, 1997). For ACS a set of non-interactive software tools, called aXe, has been developed which, on the basis of a catalogue of the objects on the associated direct image, extracts one and two-dimensional spectra (Kümmel et al., 2008). This software package was, for example, extensively used for extracting ACS Wide Field Camera slitless spectra in the Hubble Ultra Deep Field (HUDF) (Pirzkal et al., 2004). For the HLA project, we have produced a pipeline that is designed to extract spectra from large numbers of datasets in an unsupervised manner and can be tailored to particular instruments.

We chose to start the slitless spectroscopy HLA project with the NICMOS G141 grism dataset. After a brief description of the available data in the archive (Sect. 2), the basic reduction steps applied to the NICMOS images are described in Sect. 3, followed by a detailed discussion of the spectrum extraction procedure (Sec 4). In Sec 6, we show a new set of NICMOS spectrum calibration data.

Our extraction pipeline, PHLAG, is described in (Sect. 5). The extensive set of metadata that allows the archive of spectra to be served through Virtual Observatory (VO) interfaces is summarised in Sect. 7.2. The calibrated spectra were subject to quality assessment in terms of their astrometric and spectrophotometric properties and shown by internal and external comparisons to be well-calibrated (see Sect. 8).

2. NICMOS Grism Data

NICMOS, a second generation HST instrument, was installed in 1997. One of the three NICMOS cameras, NIC3 includes three grisms for slitless spectroscopy. A grism combines a grating and a prism in such a manner that it produces a dispersed spectrum

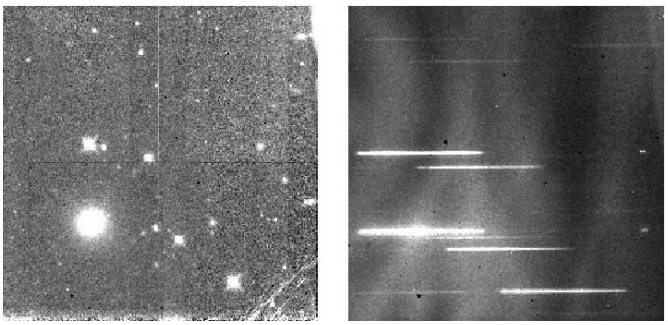

Fig. 1. Example of one matching pair of a F160W image (left, image n4k6j1a4q) and the corresponding G141 grism image (right, image n4k6j1zyq) from HLA dataset N4K6IZZCQ

at the position of the undispersed image when the grism is removed. At the nominal grism wavelength, the image through a grism will be, to first order, identical to an image without a dispersive element.

In NICMOS, the grisms are mounted on a filter wheel that can switch between any of the filters and grisms. Since there are no slit masks, spectra of all objects in the fields are produced. Depending on the location of the object, the zeroth, first or the second order spectra may be visible. The mode of operation is to first take an image with a filter that matches the bandpass of the grism followed by one or more exposures with the grism. To avoid image defects, a small or large scale dither pattern is used in a well designed observing run. Fig. 1 shows a typical pair of exposures of a filter image with a matched grism image.

The most widely used of the NICMOS grisms is G141, which covers the wavelength range from 1.10 to 1.95 μm. This wavelength range is not easily accessible from the ground, and the dark background in space makes G141 a sensitive mode to obtain spectra of all objects within the 51×51 arcsec^2 field of view of the camera. The dispersion for the first order spectrum is about 8.0 nm per pixel, which results in spectra about 105 pixels long.

In March 2002, the NICMOS Cryocooler System (NCS) was installed. With the NCS, the NICMOS detectors operate at a higher temperature and this changed the dark current as well as the quantum efficiency as a function of wavelength for each pixel. Therefore, all calibrations have to be derived separately for data taken with and without the NCS (Schultz et al., 2003).

In July 2007, the HST archive contained about 9000 NICMOS G141 observations. These data include both pointed observations towards specific targets and programmes which used NICMOS in parallel to other HST instruments. More than 50% of the NICMOS data presented in this paper are from four different HST programmes, namely HST proposal IDs 9707 (MacKenty, 1997b), 9865 (Malkan, 2004), 8082 (MacKenty, 1997a) and 10226 (Malkan, 2003). The pointing positions are almost randomly distributed on the sky and are shown in Fig. 2.

3. Image Processing and Target Selection

3.1. Associations

Each spectrum in this work is based on a collection of images, which includes both grism images and undispersed images. The images in each of these *associations* overlap and were treated collectively. The undispersed images were co-added for the creation of target lists, and the two-dimensional spectra on the grism images were rectified and co-added before the extraction of one-dimensional spectra. These steps are described in detail in this section.

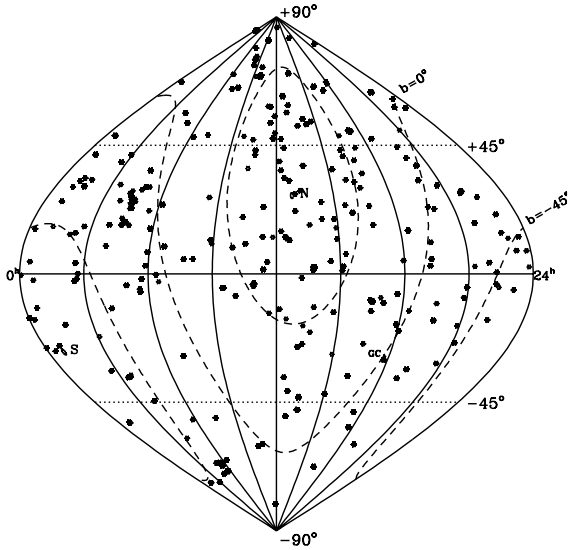

Fig. 2. Distribution of NICMOS G141 grism images on the sky.

The grouping of NICMOS datasets into associations was performed using the World Coordinate System (WCS) in the image header to compute relative pointings. To ensure accurate relative coordinates among the images in any given association, only datasets obtained with the same set of guide stars were included.

The spectral dispersion of NICMOS grism images is along the x-axis of the detector. The orientation of the grism images on the sky determines which of the spectra overlap. We therefore chose to co-add only grism images that differ in orientation by less than one degree, whereas the orientation of undispersed images was not restricted. The result of this procedure is that, for some targets, several spectra are extracted and not co-added. The reason for this decision was that different orientations of the spectra result in different contamination. Indeed, the very reason that many programmes have observed the same field with different roll angles is that it is likely that at least one of the spectra for each target is free of contamination. We leave it up to the user to pick the best of the spectra for each target.

In total 962 NICMOS G141 associations were created from 9262 grism members (up to June 2007). The grouping efficiency is therefore more than 9.5 exposures per association.

3.2. Preprossessing of NICMOS Images

To produce one-dimensional spectra of individual targets, we extracted the flux and wavelength information from NICMOS grism images and associated undispersed images that have been reduced with STScI's CALNICA reduction pipeline version 4.1.1 (Mobasher & Roye, 2004). NICMOS supports non-destructive multiple readouts, and virtually all NICMOS data are taken in this mode. The STSDAS program CALNICA (Bushouse, 1997) produces calibrated images from these raw data, the processing steps include bias correction, dark current subtraction, and computation of the count rate for each pixel using all readouts. The NICMOS images suffered from a number of peculiar anomalies, some of them are treated by CALNICA,

for others separate STSDAS tasks are available. The default parameters of CALNICA and some of the algorithms used in the auxilliary task are optimised for use with undispersed images. In the rest of the section, we describe the parameters and procedures we used for the NICMOS Grism HLA.

3.2.1. Bars Correction

Some NICMOS images suffer from pairs of bright and dark columns or rows, so-called "bars". These bars are bias-related artefacts in the form of noiseless offsets of a few DNs along a pair of columns or rows, with the pattern replicated exactly in all four image quadrants. CALNICA has a built-in procedure to detect and remove such bars. Unfortunately, the central two rows of NICMOS spectra contain more than 80% of the flux of point sources and the spectra are therefore of similar width to the bars. As a result, in crowded grism images when the spectra are well aligned with the rows of the detector, CALNICA often mistakes the peak of the spectra for bars and tries to remove them. To avoid this problem, a value of 4σ was used for the threshold in CALNICA which controls the bar detection. This compares to a default value of 3σ. It was found that this selection removes the described problem in virtually all cases.

3.2.2. Bad Pixels

Images reduced with CALNICA typically contain between 20 and 100 unflagged bad pixels, which contain significantly higher flux than their neighbours. These "hot" pixels can produce spurious emission lines in extracted spectra. The pixel spacing of the NICMOS NIC3 camera is 0.2 arcsec which undersamples the Point Spread Function (PSF). Commonly used methods to distinguish between point sources and hot or cold pixels, based on the sharpness of features on the calibrated image, are therefore not very effective with NICMOS. On the other hand, the multiple readouts provide information that can be used to recognise unreliable pixels. The following procedure was implemented to identify bad pixels. For each pixel the accumulated count rate at each readout, which is one of the outputs of CALNICA, was fitted by a line. A pixel was flagged as "bad" when either the slope of the line was different from zero by at least four times the uncertainty of the slope, or the reduced χ^2 of the fit was greater than four. Flagged pixels were ignored in the subsequent analysis.

3.2.3. Cosmic Rays

Because of the multiple readouts, the count rate of a pixel can be measured even if that pixel is hit by a cosmic ray (CR). The extra charge deposited by the CR between two readouts produces a sudden discontinuity in the count rate. CALNICA includes a procedure to identify and remove such jumps from the data. This procedure is effective for more than 90% of all CR hits. CALNICA version 4.1.1. sometimes fails and produces a bright pixel in the output images, even though it correctly recognised that this pixel is affected by a CR. A future release of CALNICA will fix this problem (Dahlen et al., 2008) but is currently not yet available. Because of the undersampling of the NICMOS images, distinguishing between stars and cosmic rays cannot be easily done on the basis of the sharpness of a flux peak alone. We therefore adopted a procedure to use both the charge built up and the sharpness to minimize the number of false detections of CR hits. Each pixel identified by CALNICA as affected by CR was checked to determine whether this pixel had an unusually high flux in the output images. Specifically, such a pixel was flagged as bad if its flux was more than 4σ above the mean of its neighbouring pixels.

3.2.4. Pedestal

NICMOS images suffer from random DC offsets in the bias level, which may be different in each of the four quadrants. The offsets are constant within each quadrant. If not removed, these offsets imprint the structure of the flatfields on the calibrated images during the flatfielding step. Since grism images are not flatfielded as images, the pedestal is visible as different constant offsets of the quadrants. There are two STSDAS tasks, pedsky and pedsub, available to remove the pedestal from the calibrated images. Neither of them works well with grism images. We therefore used the following approach to improve the pedestal in the grism images. First, for each pixel of a quadrant we normalised the count rates of the multiple readouts to the mean of the last three reads. We then computed the median count rate for each readout of all the pixels in a quadrant. We then computed the value of the DC offsets for each quadrant that results in constant median count rate. These values were then subtracted from all the pixels in the corresponding quadrant.

3.3. Co-addition of Undispersed Images

In order to create a deeper and cleaner direct image, to be used to prepare the input object catalogue needed as input to the spectral extraction, the direct images in each association were first combined. The MultiDrizzle software (Koekemoer et al., 2002; Fruchter & Hook, 2002) was used to register each input image, based on its header WCS, and to remove the small NICMOS geometric distortion using the standard cubic polynomial coefficient set. Standard default settings were used. The data quality arrays were also aligned by this step and registered output weight images created. MultiDrizzle is not yet well optimised for NICMOS and experiments revealed that a combination of inadequately precise alignment and the strong undersampling of the NIC3 camera resulted in the cores of compact objects being incorrectly flagged as cosmic rays during the "driz_cr" MultiDrizzle step. This led to significant photometric errors. To avoid this we used MultiDrizzle only up to the image combination step. The preparation of image bad pixel masks and the final image combination were performed using the *imcombine* task in IRAF.

3.4. Target Catalogue

The first step in the extraction of slitless spectra is to find objects on the undispersed image. Object parameters relevant for the extraction are the coordinates, the position angle, the object sizes, and the magnitude in a reference spectral band.

We used the SExtractor program (Bertin & Arnouts, 1996) to generate the object catalogues. The output exposure time image generated in the co-addition step was used as a weight map, which significantly reduces the number of spurious detections and improves the accuracy of the photometry.

The relevant parameters used for the extraction are given in Tab. 1. We used the "windowed" centroid (`XWIN`, `YWIN`) as the target coordinates, which is the iteratively computed first moment of the object's surface brightness after convolution with a Gaussian function with size matched to the object. This leads to very accurate centroid measurement, especially for point sources

Table 1. Relevant parameters used for the source extraction in the direct images.

Parameter	Value	Description
DETECT_MINAREA	4	Minimum number of pixels above threshold
DETECT_THRESH	3.0	Minimum threshold for detection
ANALYSIS_THRESH	3.0	Minimum threshold for analysis
DEBLEND_NTHRESH	8	Number of deblending sub-thresholds
DEBLEND_MINCONT	0.03	Minimum contrast parameter for deblending
SEEING_FWHM	0.26	Stellar FWHM in arcsec
BACK_SIZE	64	Background mesh size
BACK_FILTERSIZE	1	Background filter size

(the large majority of sources in our sample). In addition, we also measured "un-windowed" morphological major axes A, B, position angle THETA, and magnitudes (MAG_AUTO. Finally, we kept track of the success of these measurements by retaining the SExtractor FLAGS parameters (see below).

The original SExtractor catalogue was then further filtered in order to remove spurious detections:

- We removed all objects with FLAGS $\geq$ 64: that effectively removes all objects for which a severe memory overflow occurred either during deblending or during the extraction;
- We removed faint objects for which we do not expect to be able to extract a spectrum from the dispersed images (MAG_AUTO $\geq$ 23.5), or extremely elongated ones (A/B > 4), which are often simply detector artifacts or cosmic rays;
- We took objects with sizes smaller than one pixel ($\sqrt{\mathrm{AB}}$ < 1 pixel) as point-like sources, and set both their axes to 1 pixel;
- For extended objects, we computed the extension and orientation of the virtual slits from the parameters A, B, and THETA (see Appendix A);
- We discarded associations that suffered from significant crowding (> 140 sources arcmin^{-2}) since such a high source density leaves too little sky background to allow the estimation of the local background levels.
- We included external 2MASS sources in the final catalogue, as described in Sect. 3.6.

3.5. Astrometric calibration

The astrometric coordinate system of the raw NICMOS images is specified in the WCS keyswords in the image headers. Its accuracy is ultimately limited by the accuracy of the catalogue used for pointing the telescope. For earlier data, the Guide Star Catalog I (Lasker et al., 1990) was used, and since 2000 the Guide Star Catalog II (McLean et al., 2000) is used. Both catalogues are based on scans of the photographic Sky Survey plates at various epochs and bandpasses. The absolute accuracy reaches $\sim 0.3''$ over a large fraction of the sky, but errors can be as high as several arcsecs towards the edges of the scanned plates. A comparison of the astrometry in undispersed NICMOS images with more accurate data showed such large offsets in some cases.

Accurate coordinates are not only important for the final astrometric fidelity of our spectra, but also to identify potentially contaminating objects outside of the NICMOS images (see Sect. 4.9). We therefore decided to carry out an independent astrometric calibration of all our undispersed images. For that purpose, we first generated preliminary versions of the catalogues described in Sect. 3.4 with coordinates based on the WCS and used them to compute offsets to the WCS.

The offset correction applied to the WCS is the mean offset of the WCS catalogue coordinates and a global astrometric reference catalogue obtained by merging the five catalogues listed in Tab. 2. The merged astrometric catalogue includes sources from all reference catalogues and assigns to each object the coordinates of the most accurate catalogue in which it was detected. The mean offset was then calculated using a weighted average, with weights proportional to the inverse of the square intrinsic astrometric errors associated with each source.

The most critical part of the whole procedure is the matching of the sources in the astrometric catalogue with our object catalogue. This step is challenging because of the small field of view of the NIC3 camera and the relatively large number of artefacts present in the NICMOS images. The basic approach of our adopted procedure is to first use the WCS coordinates and identify the nearest neighbours in the merged reference catalogues, compute offsets to the WCS, and iterate. The scatter in the offsets of different stars is then used as an estimate of the astrometric accuracy.

Our method, however, differs in a number of ways from the simple scheme outlined above. Firstly, the initial astrometric solution assumed is not the original HST astrometry, but rather the HST astrometry corrected for the median offset computed among all the pre- or post-NCS observations. This median offset is $\sim 1.1''$ for pre-NCS observations, and $\sim 0.3''$ for the post-NCS observations (see Fig. 4).

Secondly, we used the *median* offset between the astrometric catalogue and the NICMOS object catalogue in the first two iterations, and a σ-clipped weighted average for subsequent iterations. This ensures that the initial, highly uncertain offset does not introduce any strong bias because of wrongly matched pairs.

Thirdly, if during any iteration we found more than three matched pairs, we uses a clustering analysis to remove spurious matches. Specifically, we identified clusters in the offset plane $(\mathrm{d}x, \mathrm{d}y)$. We then removed from each cluster those matches that increase the standard deviation of the offsets by more than a factor of two. This technique is more robust in removing badly matched pairs than σ-clipping.

Finally, in the case of the 2MASS catalogue, we also compared H-band magnitudes measured on the NICMOS images with those in the 2MASS catalogue and removed bad matches.

At the end of the whole process the algorithm returned the following quantities:

- the estimated offset;
- the theoretical statistical error on the offset, derived from a simple error propagation of the intrinsic catalogue accuracies listed in Tab. 2;
- in cases where at least two objects are matched, the sample variance of the offsets of the matched objects was also calculated;
- finally, the full list of matches was retained for inclusion in the metadata.

Table 2. Reference Catalogues used for Astrometric Calibration.

Catalogue	Acronym	Reference	Accuracy
USNO CCD Astrograph Catalog	UCAC2	Zacharias et al. (2004)	0.06 arcsec
Two Micron All Sky Survey	2MASS	Skrutskie et al. (2006)	0.10 arcsec
Sloan Digital Sky Survey	SDSS-DR5	Adelman-McCarthy et al. (2007)	0.15 arcsec
USNO-B1.0 catalog	USN0	Monet et al. (2003)	0.20 arcsec
Guide Star Catalog II	GSC2.3.2	McLean et al. (2000)	0.30 arcsec

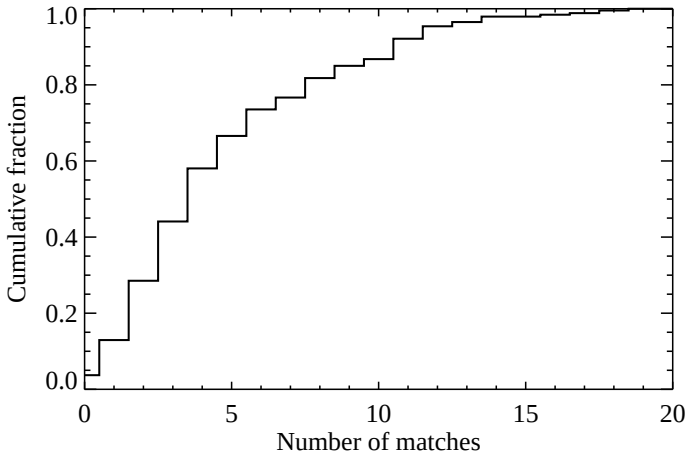

Fig. 3. Cumulative distribution of number of astrometric calibration stars used in each field.

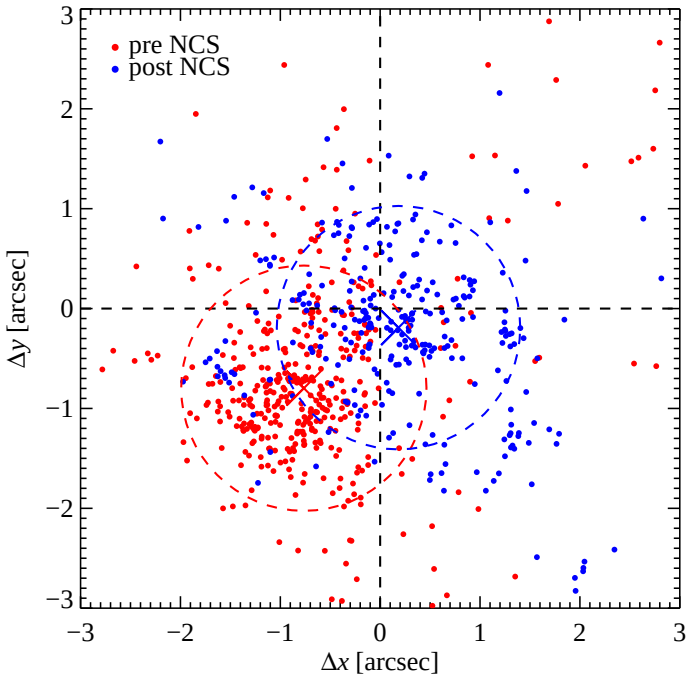

Fig. 4. Difference between objects coordinates computed from the WCS in the image header and coordinates listed in astrometric reference catalogues. The crosses mark the mean of the distributions before and after the installation of the NCS, and the circles make the 1σ scatter around the mean. The systematic error in the WCS coordinates is significant larger for the pre-NCS data.

In total, about 96% of our released spectra have corrected coordinates, and for 91% of these we were able to match two or more astrometric sources The distribution of the number of matches is shown in Fig. 3 The computed offsets are shown in Fig. 4. In Fig. 5 we compare the distributions of the estimated astrometric uncertainty (i.e. the square root of the sample variance of offsets) before and after applying these offsets. The median accuracy of the astrometry improves from about 1.24 arcsec to about 0.27 arcsec.

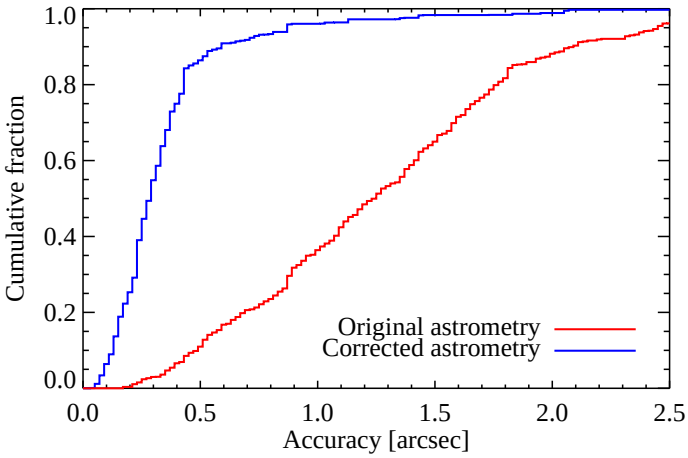

Fig. 5. Cumulative distribution of estimated astrometric accuracies of the object coordinates before and after the astrometric correction.

3.6. External 2MASS sources

NICMOS grism images can include partial spectra of bright objects that are located outside of the field of view (FOV) but close enough to the edge of the image so that the grism still deflects part of the first order spectrum onto the detector. The zeroth order spectra of such objects can be fully included in the grism image without a corresponding counterpart on the undispersed image. Such partial spectra might contaminate other spectra, and the zeroth order spectra might be mistaken for bright emission lines. We used the 2MASS catalogue to identify bright sources close to each grism image that might create such spurious spectra. After the astrometric correction described in Sect. 3.5, the coordinates of the images are known well enough to predict the location of such spectra accurately enough to take them into account for contamination estimation. By visual inspection of the images we determined that the maximum distance of sources from the image edges that still produces a spectrum is 13 arcsec at the left image edge, and 3 arcsec at the right image edge.

Although it is in principle possible to extract spectra of the external sources, in the current release we did not attempt to do so, but only used the external sources for the estimate of the contamination see Sect. 4.9.

4. Extraction of NICMOS Spectra

4.1. Methods

One-dimensional integrated first order spectra of objects were extracted from grism data with the help of the target catalogues.

Whilst second order spectra are visible on NICMOS grism images for bright targets, in most cases only small parts of these spectra are usable because they are not completely contained within the NICMOS detector and usually suffer from significant contamination. We therefore chose to limit our project to the first order spectra.

The essential extraction steps were to remove the local background close to the spectrum, use the position of the object on the undispersed image to obtain a wavelength scale, and then add up the flux values for each wavelength bin. The wavelength dependent response of each pixel was taken into account during this extraction.

We used a modified version of the extraction software aXe (Kümmel et al., 2008) version 3.7 for that purpose. aXe was designed to treat slitless spectroscopic data and includes the transfer of source positions derived on undispersed images onto the slitless images; the application of wavelength dependent flat fields, the co-addition and extraction of spectra and the estimation of the contamination. We now describe the modifications and additions that were made to the basic aXe package, most of them address NICMOS specific properties of the data.

4.2. Tracing of spectra

Spectra on G141 grism images are closely aligned with the rows of the detector. The angle between the dispersion direction and the detector rows depends on the exact orientation of the grisms. The grism is mounted on a rotating filter wheel, which lead to slight differences in the positioning of the grisms each time it is rotated into place. Typical values for the angle between the dispersion direction and the x-axis of the detector ranged from 0 to 2°. We therefore determined this angle separately for each grism observation.

To ensure an identical extraction around the variable traces for the entire data set, the spectral trace was measured on each grism image individually. This was done by selecting the brightest three point-like objects on each image and determining the trace solution for each object individually. The individual solutions were then combined using the signal-to-noise ratio of the spectrum as weight to determine the single trace solution for that grism image.

4.3. Background Subtraction

The NICMOS grism images are not flatfielded in the standard pipeline, because the local quantum efficiency is wavelength dependent. Spatial variations in the throughput were therefore taken into account during the extraction process for each target spectrum, after the wavelength calibration. An imprint of the flatfield convolved with the spectrum of the sky background in the grism passband is therefore visible in calibrated grism images. This background has to be *subtracted* before the flux in each wavelength bin is summed. In addition to the imprint of the flatfield, a region of enhanced dark current was visible in the lower right-hand part of all NICMOS images. This enhanced dark current, which varies depending on the positioning of the field offset mirror (FOM), is also an additive component that has to be subtracted.

G141 background images were constructed from grism images that contained no or few visible spectra, using about twenty different images for each prepared background. The median of the scaled images was subtracted from each individual image to isolate the contribution of the FOM to the background. In this manner, two different background images were prepared for different time periods, one presenting the flatfield imprint on the background, and the other the enhanced dark current. Scaled versions of these backgrounds were subtracted from each grism image.

The scaling of the subtracted background was a crucial parameter for the flatness of the resulting image and therefore the quality of the extracted spectra. Because of the uncertainties in the overall bias level and dark current and the changing sky brightness, the measured background level was not a good predictor of the structure seen in the background. The scaling for the subtraction was therefore determined by finding the least-squares solution for a line fit to the pixel values in the background versus the ones in the grism images. Only pixels that do not contain any spectrum were used in this procedure.

For most data sets, this procedure succeeded in removing most variations in the background. To further improve the quality of the extraction, a local background was subtracted for each extracted spectrum. This local background is a linear fit to the region around each spectrum.

4.4. Virtual Slits

The slit in a longslit spectrograph selects a part of the sky, and light from that area is then dispersed and binned in wavelength. In slitless spectrographs, the light that reaches the final wavelength bins is only limited by the extent of the targets. The light distribution of target objects therefore defines the lines of constant wavelength on the dispersed image, and the effective resolving power. The size, position and orientation of the lines of constant wavelength are used to extract one-dimensional spectra from the two-dimensional image.

Target objects can be of an arbitrary shape that may depend on wavelength. For the extraction of spectra, we made two approximations. Firstly, we neglected any wavelength dependence of the object shape and derived all shape parameters from the undispersed image. Secondly, we treated all objects as ellipses, i.e. assumed that the isophotes were elliptical. The spectral resolution of the spectra extracted with these approximation was slightly lower than an optimal extraction that uses knowledge of the exact shape of objects as a function of wavelength. For these approximations, we computed the extraction direction that optimizes the spectral resolution of the spectra. The details are given in Appendix A.

4.5. Adjustment of Wavelength Scale

As a result of the filter wheel non-repeatability, the spectra on any given exposure can be shifted by up to a few pixels relative to the expected wavelength zero point derived from the undispersed image. The wavelength zeropoint was therefore re-adjusted for each exposure by extracting the brightest few (~4) spectra and cross-correlating the characteristic spectral feature (i.e. the drop of sensitivity) at the red end of all spectra with a fiducial template. This template was prepared from the sensitivity curves shown in Fig. 8. In this way any wavelength zero-point shifts introduced by the non-repeatability of the positioning of the grism wheel were taken out with a precision of about ±0.2 pixel or ±16 Å. Potential wavelength calibration changes as a function of position within the field of view were not taken into account.

Table 3. Tabulated values for the linearity correction b

$\lambda[\mu]$ m	.825	.875	.925	.975	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
b	.069	.057	.052	.050	.049	.048	.041	.023	.013	.008	.004	0

4.6. Pixel Response Function

The NICMOS detectors show significant sensitivity variations across individual pixels (Storrs et al., 1999; Xu & Mobasher, 2003). These variations, together with the large pixel size and the small trace angles (see Sec 4.2), result in a modulation of the extracted spectrum with a wave-like pattern for point-like objects. The amplitude of this pattern is about 4%. To correct for this effect the exact location of the trace must be known to a fraction of the pixel size.

The subpixel accuracy for the trace location within a spectrum cannot be achieved with the fits described in Sec 4.2. A sensitive estimator of the trace location is the quantity $f = v_1/v_2$ with the brightest pixel value v_1 and the second brightest pixel value v_2, respectively. For each point-like object, the values f along the spectrum were determined. The position where this function peaks marks the wavelength value at which the spectrum is located at the centre of a row. For raw spectra with sufficient signal-to-noise to determine the peaks were multiplied by a correction function $f_{\rm pr}$:

$$f_{\rm pr} = 1 + 0.04\sin(2\pi y), \quad (1)$$

where y is the position of the spectrum on the detector determined with subpixel accuracy as described above.

4.7. Non-Linearity Correction

The NICMOS detector responds to different infalling photon count rates in a non-linear manner. Even if the total number of photons is identical, the detector will report a different flux if the number of photons per unit time differs. This effect was described in Bohlin, Riess & de Jong (2006). Following Bohlin, Lindler & Riess (2005), we corrected point-like objects for the detected photon count rates c according to

$$c = c_{\rm obs}/[1 - 2b(\lambda) + b(\lambda)\log_{10}(c_{\rm obs})], \quad (2)$$

where c_{obs} is the count rate derived from the grism image, and the values $b(\lambda)$ are given in Table 3.

4.8. Flux Calibration

Flux calibration of point sources was done by multiplying the count rate of extracted spectra by the sensitivity curve (see Fig. 8) derived from flux standards. Flux standards are stars and thus point-like sources, and the spectral resolution is determined by the instrument setup. For extended objects, however, the spectral resolution is degraded by the extent of the object in the dispersion direction. This smoothing of the spectra reduces the amplitude of structure in the sensitivity curve and has to be taken into account for the flux calibration of extended objects.

The flux calibration of extended sources was carried out by multiplying the count rate by a smoothed version of the point source sensitivity function. The width of the Gaussian smoothing kernel was

$$\sigma = fr\sqrt{s^2 - p^2}, \quad (3)$$

where s is the FWHM of the object in the dispersion direction, p is the FWHM of the point spread function, and r is the dispersion in Å per pixel. The quantity f is a correction factor that was found empirically by experiment using extended source spectra; the adopted value was 0.65.

The procedure was tested both on simulations of Gaussian objects using the aXe simulation package aXeSIM (Kümmel, Kuntschner & Walsh, 2007), and on real extracted NICMOS spectra. The tests showed that, in most cases, the procedure successfully removed the imprinted structure resulting from the changes in sensitivity with wavelength that is visible in spectra calibrated with the nominal point source sensitivity curve.

4.9. Contamination

One of the disadvantages of slitless spectroscopy is that spectra of science targets may overlap with the spectra of random objects in the field that are not masked by any slit mechanism. Only isolated objects are not affected by this sort of contamination. The level of contamination depends on the separation and relative flux level of the objects. Spectra that are contaminated may still be useful if the expected flux from the contaminator is significantly smaller than the flux in the spectrum of interest. Estimating the contamination level for all spectra is essential to exploit them scientifically.

Contamination levels were estimated by aXe assuming flat contaminating spectra (Kümmel et al., 2008). We used this feature to estimate the contamination both from objects in the source catalogue and from bright 2MASS sources (see Sect. 3.6) that were close enough to the target object so that their spectra might overlap. The contamination level was estimated in the following manner: The shape of spectra in the cross dispersion direction was assumed to be Gaussian with a dispersion σ equal to the object size listed in the target catalogue. The spectrum of each potential contaminator was assumed to be flat in f_λ with an integrated magnitude identical to the one measured on the undispersed image. With these assumptions, two-dimensional spectra for all potentially contaminating objects were produced. Finally, a spectrum at the position of the target was extracted in exactly the same manner as the spectrum from the original grism image. The resulting spectrum was then used as an estimate of the contaminating flux.

4.10. Co-addition of Spectra

The final step in the extraction of spectra is to sum up all flux contributions for a given wavelength bin. The approach used by aXe is to rectify the two dimensional spectra onto a predefined grid using the drizzle approach (Koekemoer et al., 2002). In our modified version of aXe, each spectrum of a target is first drizzled to the same grid. We then use the IRAF task *imcombine* with 4σ rejection to combine the two-dimensional spectra. One-dimensional spectra are finally produced by summing up all pixels that correspond to the same wavelength.

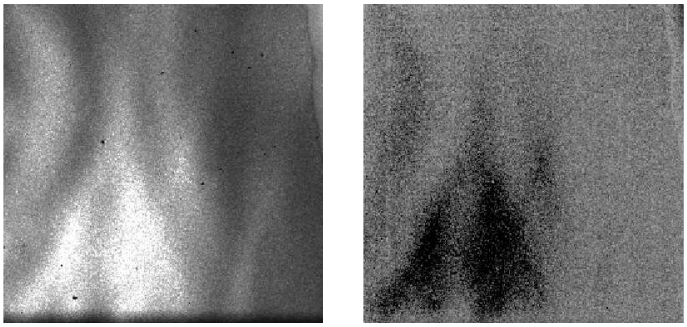

Fig. 6. The zeroth order (wavelength independent) plane of the post-NCS flat field cube (left) and the first order (wavelength dependent) plane (right).

5. The Pipeline for the Hubble Legacy Archive Grism data (PHLAG)

The whole extraction procedure from the retrieval of the input data to the extraction of the calibrated spectra was carried out in a single Python script, called **P**ipeline for the **H**ubble **L**egacy **A**rchive **G**rism data (PHLAG). PHLAG works with the NICMOS calibrated images and calls the necessary external packages in a series of processing steps. The steps carried out are as follows:

1. **Data preparation:** The data are prepared for the pipeline reduction. The direct images are grouped according to the filter. Pairs, consisting of one direct image and one slitless image with small positional offsets, are composed.
2. **Image combination:** The direct images are rectified and co-added as described in Sect. 3.3.
3. **Object detection:** The object catalogues are extracted from the undispersed image (Sect. 3.4).
4. **Spectral extraction:** The extraction of one and two-dimensional spectra is performed using the aXe software package with the modifications described in Sect. 4.
5. **Metadata:** The spectra are prepared for ingestion into the database. The metadata are collected or derived (see Sect. 7.2). Associated products such as stamps and cutout-images are created.

6. Calibration of NICMOS Spectra

6.1. Flatfield Cube

The quantum efficiency of each pixel on the detector is a function of wavelength, and this function is different for each pixel. For slitless spectroscopic images, the usual step of correcting the quantum efficiency variations by flatfielding has to be replaced by a wavelength dependent correction. Such a correction requires a three-dimensional flatfield cube with detector coordinates and wavelength as the three axis.

The flatfield cube was derived from the NICMOS narrow band flat fields, separately for the pre-NCS and post-NCS periods. Tab. 4 lists the narrow band flats that were used. For each pixel, we fitted a 5th order polynomial to the value of the pixel in each flatfield. Fig. 6 illustrates the zeroth order (wavelength independent) and first order (wavelength dependent) terms of the flat field cube. The mean value of the rms on the polynomial fit for the post-NCS flat field cube is 0.15%.

Table 5. Nebular emission lines used for wavelength calibration.

λ(Å)	Species
10830.290	HeI
10938.095	HI
11164.403	[Fe II]
11305.854	[SI]
11969.059	HeI
12527.506	HeI
12818.08	HI
14072.70	[Fe II]
14706.272	[Fe II]
16109.313	HI
16407.192	HI
16806.520	HI
17362.108	HI
18174.121	HI
18751.01	HI

Table 6. Dispersion solutions for G141 spectra.

Spectrum order	A_0(Å)	A_1(Å/pix)
1	13962.8	80.19
2	7141.7	40.54

6.2. Wavelength Calibration

In-orbit wavelength calibration for G141 spectra was established by observations of spectra of planetary nebulae (PNe). Two compact PNe were observed – Hubble 12 (Hb 12) and Vy2-2, although both are resolved by NIC3. Fig. 7 shows the one-dimensional extracted spectrum of Hb 12 with the identified emission lines indicated. By matching the observed positions of the lines, fitted by Gaussians, with known wavelengths, the G141 dispersion solution was established. Vy2-2 was observed in 1997 as part of the early calibration programme and Hb 12 in 2002 in the post-NCS era. For Hb 12 fifteen spectra were analysed for first order dispersion solution and eight for the second order spectrum; for Vy2-2 only three spectra were available to analyse, and were only used to verify the solution derived from Hb 12. Within the errors no difference was found between the wavelength solutions for both targets, indicating no measurable change of the dispersion pre- and post-NCS.

Up to fifteen emission lines were detectable in the first order spectra and seven for the second order; the lines and their identifications are listed in Tab. 5. First and second order polynomial fits were made to the variations of the pixel position with wavelength. The second order terms are very small, typically 2×10^4 times smaller than the first order fit coefficient; they were neglected and only a linear solution adopted. The sampling of any spatial dependence of the wavelength solutions was poor and single average values for the fit coefficients were adopted for the whole detector. Tab. 6 lists the first and second spectral order dispersion solutions in the form $\lambda(\text{Å}) = A_0 + A_1 x_0$, where x_0 is the pixel offset along the trace from the position of the direct object.

6.3. Sensitivity Calibration

The sensitivity curve needed for the flux calibration of the spectra (Sect. 4.8) was derived from observations of the flux standards GD153, G191B2B, and P330E. Tab. 7 lists the observations used to derive the final sensitivity calibration. Spectra were

Table 4. Flat field cube fitting for Pre- and Post-NCS flat fields.

Filter	Pre-NCS Flat	Post-NCS Flat
F108N	i191346kn_flt.fits	3_F108N_STEP16_1On-AllOff_sflt.fits
F113N	i191346mn_flt.fits	3_F113N_STEP16_1On-AllOff_sflt.fits
F164N	i191346pn_flt.fits	3_F164N_STEP16_1On-AllOff_sflt.fits
F166N	i191346qn_flt.fits	3_F166N_STEP16_1On-AllOff_sflt.fits
F187N	i191346sn_flt.fits	3_F187N_STEP16_1On-AllOff_sflt.fits
F190N	i191346tn_flt.fits	3_F190N_STEP16_1On-AllOff_sflt.fits
F196N	i1913470n_flt.fits	3_F196N_STEP16_1On-AllOff_sflt.fits
F200N	i1913471n_flt.fits	3_F200N_STEP16_1On-AllOff_sflt.fits
F212N	i1913472n_flt.fits	3_F212N_STEP16_1On-AllOff_sflt.fits
F215N	i1913473n_flt.fits	3_F215N_STEP16_1On-AllOff_sflt.fits
F240M	i1913475n_flt.fits	3_F240M_STEP2_1On-AllOff_sflt.fits

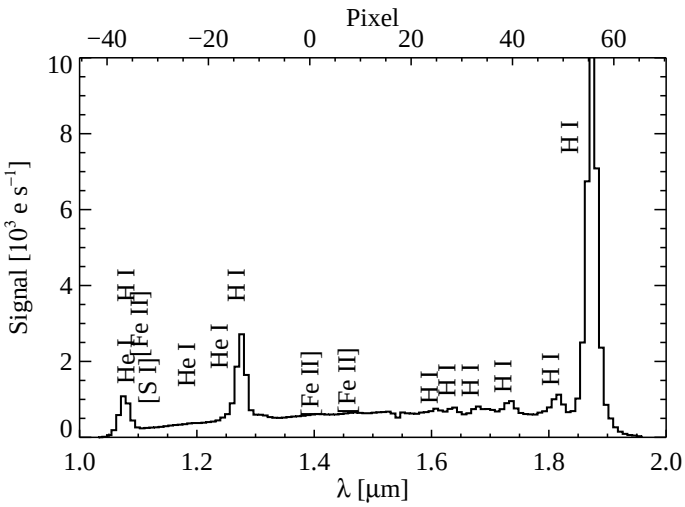

Fig. 7. NICMOS G141 spectrum of the planetary nebula Hb 12 showing the emission line identifications. The upper axis shows the offset in pixels from the position of the direct image.

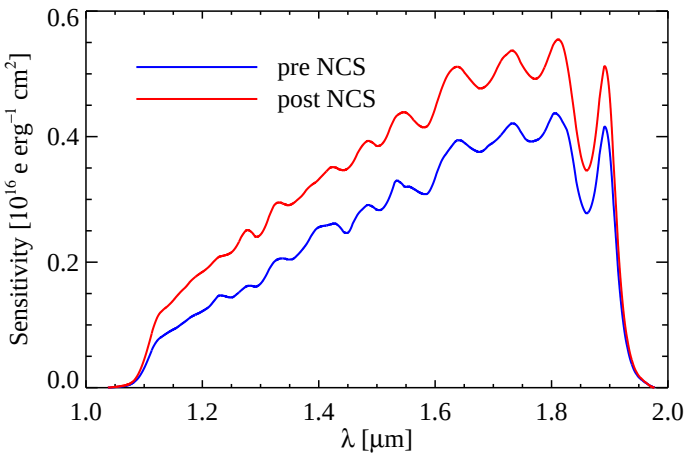

Fig. 8. NICMOS sensitivity curves as a function of wavelength for the G141 grism for pre-NCS and post-NCS data sets.

extracted from those observations and co-added for each star separately for pre-NCS and post-NCS observations. The sensitivity curve was then computed by dividing the count rates of these spectra by tabulated standard fluxes for each star. The standard fluxes were taken from the HST CALSPEC library and are based on Hubeny NLTE models (Bohlin, 2003) in the case of G191B2B and GD153 (g191b2b_mod_004.fits, gd153_mod_004.fits) and a combined STIS, NICMOS spectrum in the case of P330E (p330e_stisnic_001.fits, Bohlin, 2007).

The average flux calibrations agreed to about 3% between the three standard stars and we did not detect significant variations as a function of time of observation. However, variations as a function of position within the field of view indicated a scatter of about 8%. The field of view coverage with flux standard stars was not sufficient to allow a field-dependent sensitivity calibration. Overall the absolute flux calibration for post-NCS observations was expected to be accurate to better than 10%, while for pre-NCS data it was difficult to assess the overall reliability due to a lack of suitable standard star observations. The sensitivity files used in the pipeline reduction carry a random error of only 1-3% that was propagated into the final spectra. In addition, there was an uncertainty of about ~10% in the overall normalisation of the flux. The final sensitivity curves derived for pre- and post-NCS periods are shown in Fig. 8.

7. Data Products and Distribution

7.1. HLA Portal

HLA data are distributed by both ST-ECF and STScI. There are three main ways to search, browse, and access the NICMOS spectra:

1. Archive Query Interface. The HLA archive can be searched online (`http://archive.eso.org/wdb/wdb/hla/product_science/form`) and constraints on the target (e.g. the target name), the data properties (e.g. effective exposure time), the source properties (e.g. the magnitude) and the data quality (e.g. the signal-to-noise ratio) can be placed. The detailed result pages show the preview of the data as well as all available metadata.
2. HLA archive at STScI. The ST-ECF HLA data has been also integrated into the HLA interface of the Space Telescope Science Institute (`http://hla.stsci.edu`). A subset of the parameters of the ST-ECF HLA interface are shown and can be queried.
3. Virtual Observatory. We provide fully automated access to the HLA metadata and data via Virtual Observatory (VO) standards. A Simple Spectrum Access Protocol (SSAP) server has been established (`http://www.stecf.org/hla-vo`). It serves VOTables in V1.1 format, which contain, in addition to the standard metadata, information about the footprints of the equivalent slits of the grism spectra. Our SSAP server has been tested with ESO's archive browser VirGO (`http://archive.eso.org/cms/`

Table 7. List of flux standard stars

Star	Association	PROP ID	PROP PI	pre/post NCS
GD153	N94A02E6Q	10385	Bohlin	post NCS
GD153	N8U402MVQ	9998	Bohlin	post NCS
GD153	N9U203NOQ	11064	Bohlin	post NCS
G191B2B	N9U201M3Q	11064	Bohlin	post NCS
G191B2B	N8U405TWQ	9998	Bohlin	post NCS
G191B2B	N94A03GDQ	10385	Bohlin	post NCS
P330E	N8U406VZQ	9998	Bohlin	post NCS
P330E	N8BR01ICQ	8991	Thompson	post NCS
P330E	N9U212P8Q	11064	Bohlin	post NCS
G191B2B	N4IT01NUQ	7696	Calzetti	pre NCS
P330E	N4VD01OYQ	7959	Calzetti	pre NCS

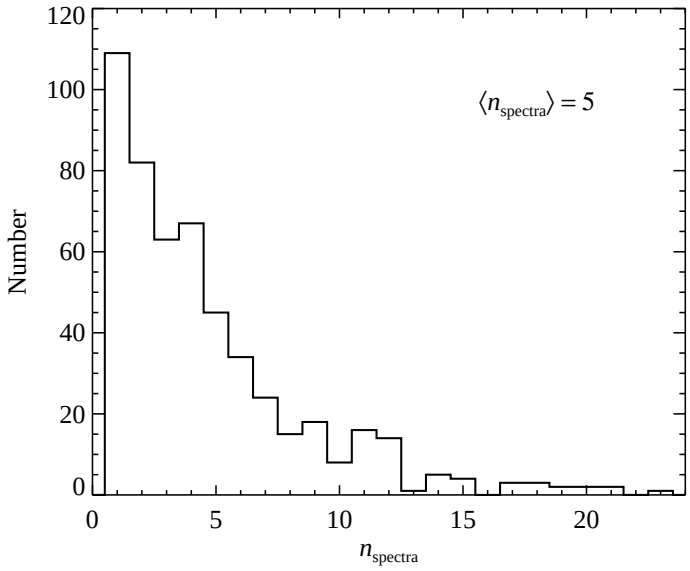

Fig. 9. Distribution of the number of extracted spectra from HLA datasets.

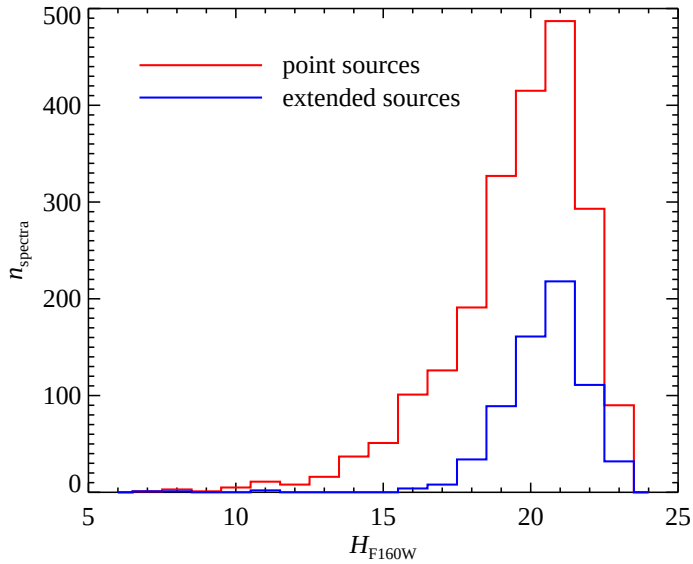

Fig. 10. Magnitude distribution of the targets with spectra in the HLA release. The blue histogram are the extended sources.

virgo/) as well as with SPLAT (http://star-www.dur.ac.uk/~pdraper/splat/splat-vo/, also available at http://starlink.jach.hawaii.edu/) and VOSpec (http://esavo.esa.int/vospecapp).

7.2. Distributed Files

The spectra are distributed as sets of FITS files, which include two dimensional cutouts of each target from the rectified and calibrated grism images, a cutout of the target from the undispersed filter image and a one-dimensional extracted spectrum.

The two-dimensional grism stamp images and the direct image cutouts are multi-extension FITS files. The one-dimensional spectrum follows the data formatting specified for FITS serialisation by the IVOA Spectral Data Model version 1.01 (McDowell & Tody, 2007). Each data point contains the wavelength in Å, the count rate in electrons per second and the flux expressed in physical units along with associated errors and an estimate of the contaminating flux. Additional metadata include keywords to describe the contamination, the orientation of the dispersion direction on the sky and footprints.

8. Results

8.1. Data Release 1

Data were released on Feb 8, 2008. This Data Release 1 (DR1) includes a total of 2470 extracted from 520 associations.

The H-band magnitudes computed from the undispersed images of the targets range from 7 to 23.5 for point sources, and from 16 to 23.5 for extended sources as identified by the SExtractor program (see Fig. 10). Only spectra with measured signal-to-noise ratios larger than three were included in the release, and typical measured signal-to-noise ratios are about 50 for objects brighter than H≈17 (see Fig. 11).

8.2. Completeness

Whether the spectrum of a particular target is extracted depends, among other things, on the detection of the target on the undispersed image, the relative location and flux levels of other spectra on the grism image, the location of residual image defects and the location of the spectrum relative to the image edges. The number of spectra extracted at each pointing therefore varies widely, and was typically about five. The distribution of the number of spectra extracted from each association is shown in Fig. 9.

To assess the completeness of the sample quantitatively, we have compared the DR1 catalogue with the survey by McCarthy

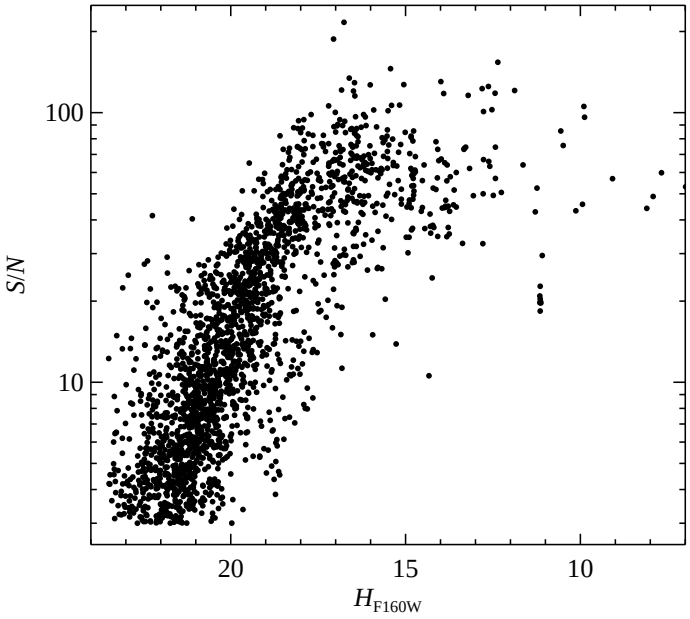

Fig. 11. Measured signal-to-noise of the HLA spectra as a function of magnitude.

et al. (1999). These authors searched two-dimensional NICMOS grism images for emission line objects and found a total of 33 candidates. Most of the emission lines are Hα at redshifts between 0.75 and 1.9. Such a sample of galaxies is dominated by closely interacting systems and, because of contamination issues, our procedure selects against such objects. We found that a total of 6 of the McCarthy et al. candidates are included in DR1. We therefore consider 20% to be a lower limit for the completeness of our catalogue.

The HLA includes any spectrum suitable for extraction, no attempt has been made to classify or match the targets with any catalogue. Some of the fields have been observed several times with different roll angle. These spectra were not co-added and therefore several spectra might be available for one target. In total, the number of unique targets in the release is 1923.

8.3. Photometric accuracy of the undispersed images

The photometric accuracy of the undispersed direct images was assessed by comparing the magnitudes obtained in the F160W and F110W bands with the 2MASS H and J band magnitudes of matched objects. When doing this comparison, we used aperture photometry with an aperture of 11 pixels, and applied a finite-aperture correction as provided by the NICMOS Data Handbook (Mobasher & Roye, 2004); in addition, we converted the 2MASS magnitudes into the AB magnitude system.

The results obtained are shown in Fig. 12: a good agreement is obtained over the whole magnitude range, with negligible bias; the relatively large scatter observed, ~ 0.16 mag, is probably due to the photometric uncertainties present in the 2MASS catalogue for faint sources. Similar results were obtained for the F110W magnitudes.

8.4. Photometric accuracy of the spectra

Quality control of the photometric calibration of the one-dimensional spectra was performed by comparing the integrated flux, as measured from the end-product spectra, with the magnitude derived from the direct images. The following procedure was applied:

- all spectra that had associated direct images in the F160W filter were selected;
- for the selected objects, only spectra which were considered complete up to the boundary of the total throughput of the F160W/NIC3 configuration (see Fig. 13), were selected. In

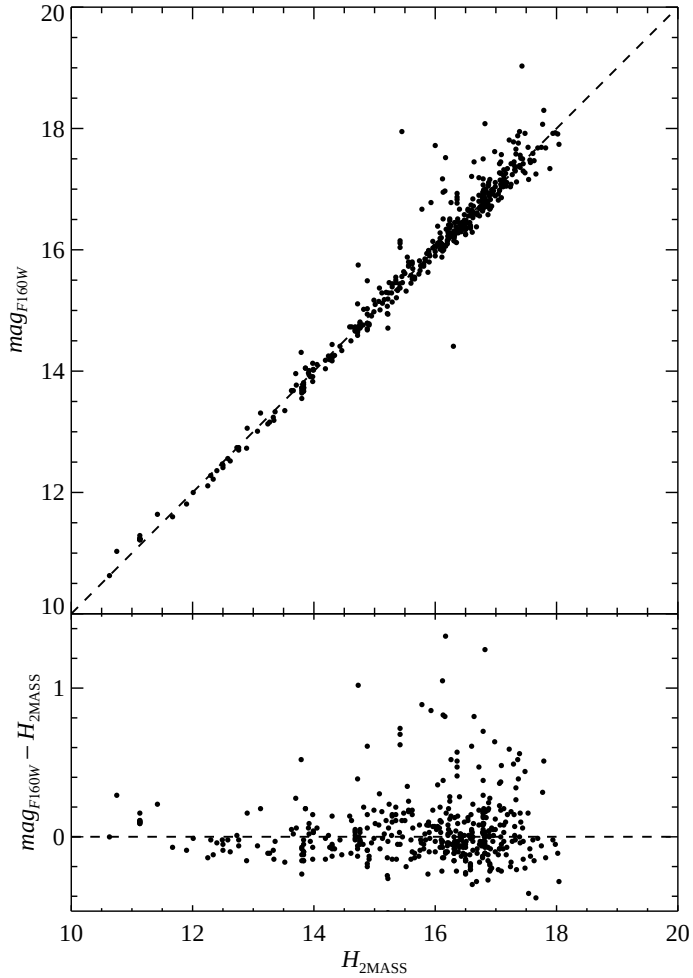

Fig. 12. Top. Direct image F160W magnitude, as a function of the 2MASS H magnitude. **Bottom.** Difference of the two magnitudes as a function of the 2MASS magnitude .

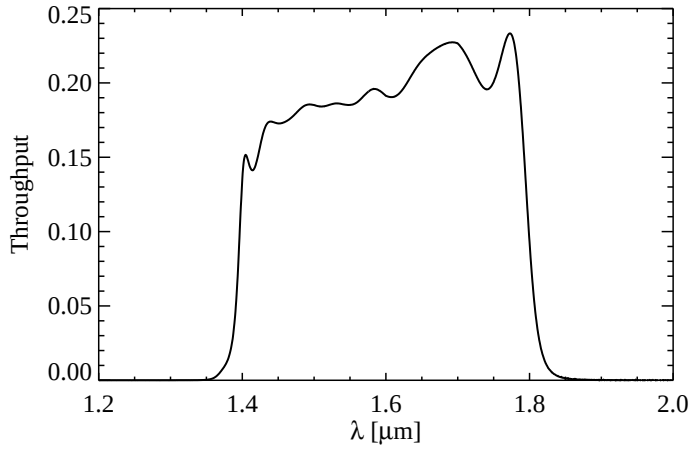

Fig. 13. Total throughput of the F160W filter for the NICMOS/NIC3 camera.

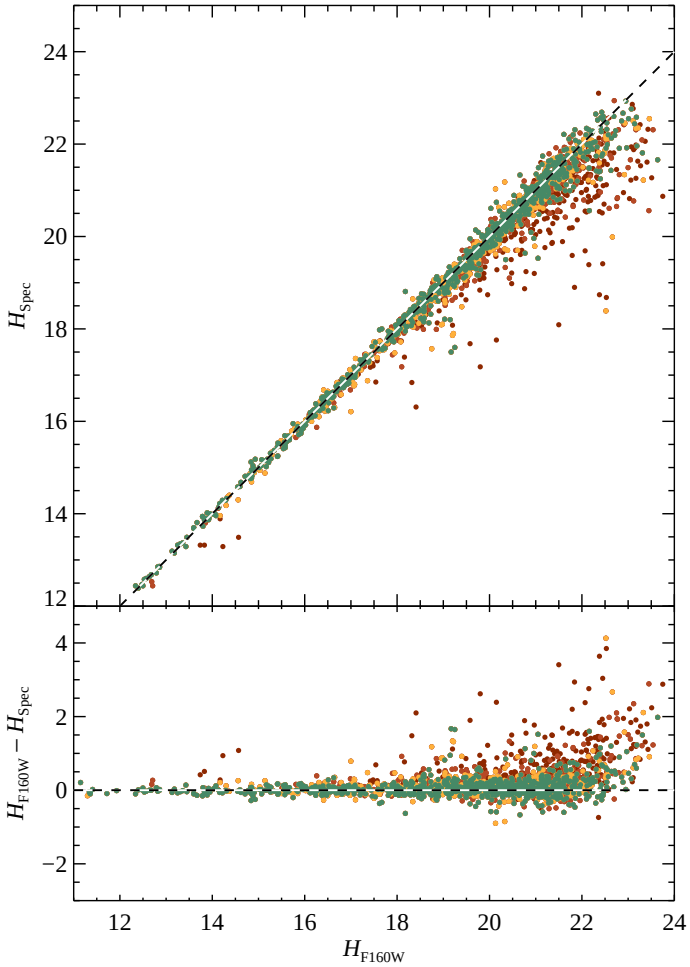

Fig. 14. Top. The "spectroscopic" magnitude, i.e. the magnitude derived from the spectrum of the object, is shown as a function of the direct image F160W magnitude. **Bottom.** The difference of the two magnitudes is plotted as a function of the direct image magnitude. For both plots, the colour of the points indicates the level of contamination, coded according to the difference Δm in the magnitude of the target and that of the contaminator. The colour codes are as follows: green: $\Delta m > 8^{mag}$, yellow: $4^{mag} < \Delta m < 8^{mag}$, orange: $2^{mag} < \Delta m < 4^{mag}$ and red: $0^{mag} < \Delta m < 2^{mag}$.

particular, all spectra that did not contain data within the wavelength range defined by throughput > 0.005 were discarded;

– the flux of these objects was then integrated and compared with either the direct image magnitudes or, when available, with the 2MASS H band magnitudes. At the same time, the error on the magnitude due to flux contamination from other sources was also evaluated using data provided in the output products.

The results obtained for the comparison with the F160W direct images are shown in Fig. 14. A very good linear response is obtained over a large range of magnitudes for both pre- and post-NCS data. The effects of spectrum contamination are visible in the lower-right part of the graph, which is occupied mainly by sources with high levels of contamination, as directly evaluated from the spectra. Overall the agreement is good, with a median offset of 0.16 magnitudes for pre-NCS data, and 0.23 for the post-NCS case. The scatter is in both cases ~ 0.5 magnitudes.

A few "catastrophic" cases of sources with apparent no-contamination but still large magnitude difference are however

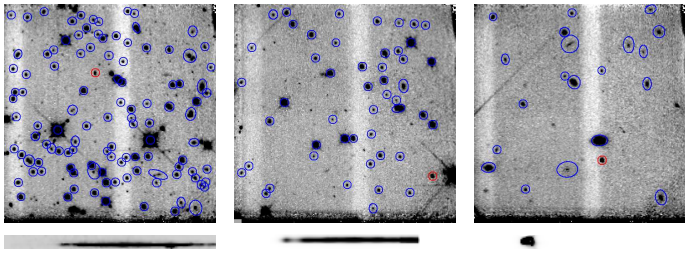

Fig. 15. Examples of catastrophic discrepancies selected from the outliers of Fig. 14. Each panel shows the direct image associated to the source for which an inaccurate spectroscopic magnitude was obtained (marked in red); other objects in the field used for the contamination calculation are marked with blue ellipses. The corresponding cutouts from the grism image are shown below the undispersed images. **Left.** Source in a crowded field, close to an undetected source to the right. **Middle.** Source close to bright star on the boundary. **Right.** Source contaminated by a bright star outside the field (note the spike on the centre-left of the image).

evident in this plot. These cases were investigated further and simple explanations could be found for almost all of them. In general, it was found that large discrepancies were associated with: (i) crowded fields, for which the contamination model might be inaccurate; (ii) objects close to bright sources not included in the analysis; (iii) objects close to extended, diffuse sources; or (iv) objects contaminated by sources outside the field of view of the direct image. Three examples of such cases are shown in Fig. 15. For each case, the undispersed image with the marked problematic target is shown along with the corresponding spectrum cutout. In the left and centre panels, a bright undetected source produces significant contamination in the extracted spectra. In both cases, the contaminating source dominates the flux. In the right panel, the zeroth order of a very bright source outside the field produces a bright region in the extracted spectrum.

8.5. Noise

All extracted spectra include a predicted flux uncertainty in each wavelength bin based on error propagation from the error array in the NICMOS images through to the final spectra. Fig. 16 compares these predicted uncertainties with noise measurements on the final spectra using the DER_SNR algorithm (Stoehr et al., 2007).

The SNR values are in good agreement with quite small scatter at the high flux end. At the faint end the scatter is larger and there is a slight but systematic trend for the DER_SNR values to be higher than their predicted value. This might be due to a component, e.g. the noise in the background, which was not included in our noise model, or reflect some bias in the measurement of the noise.

8.6. Comparison of Spectra with NICMOSlook Extractions

Most previous extractions of NICMOS grism spectra have used interactive tools, such as the NICMOSlook program (Freudling, 1997). NICMOSlook is a highly specialised tool that provides a large number of options and parameters that can be varied to optimise the extraction. The main advantages of using an interactive program is that the background region can be adjusted for each individual spectrum and that the subpixel accuracy of the trace necessary for the pixel-response correction can be visually checked and fitting parameters can be modified. In addi-

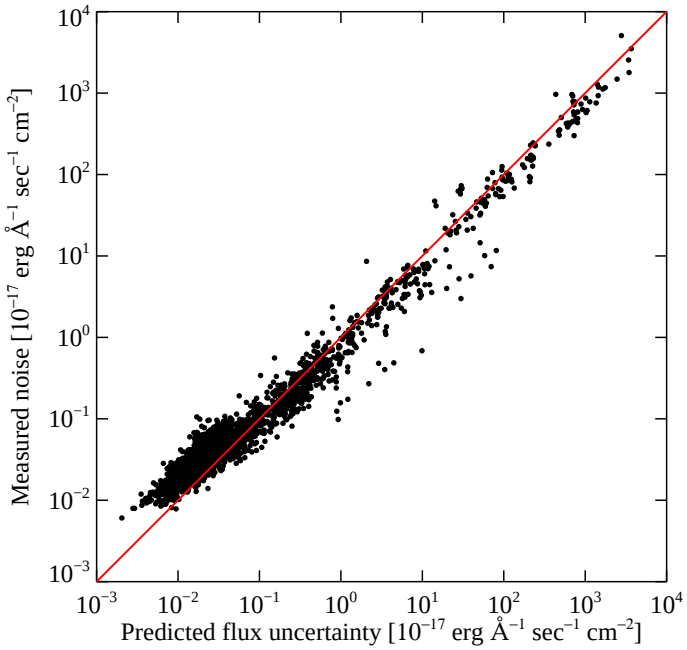

Fig. 16. Noise measurements from the spectrum using the DER_SNR algorithm as a function the noise estimate obtained from the NICMOS error array using error propagation.

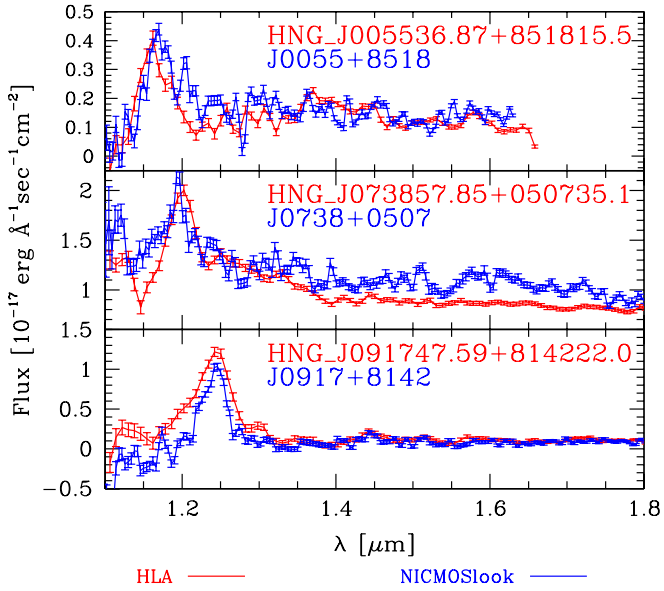

Fig. 17. Comparison of HLA spectra with extractions using the NICMOSlook program. All three spectra are Hα emitting galaxies discovered by McCarthy et al. (1999). The blue curves are NICMOSlook extractions, whereas the red curves are the HLA extractions. The blue labels are the McCarthy et al.'s names.

tion, NICMOSlook's optional deblending of overlapping spectra results in a more reliable estimate of contamination levels, and is accurate enough to recover heavily contaminated spectra. The combination of optimising the background regions and deblending of overlapping spectra allows NICMOSlook to extract more spectra in crowded regions than is possible with the automatic procedure used in this work. For example, McCarthy et al. (1999) used NICMOSlook to extract the larger sample of Hα emitters mentioned above.

For those spectra that are included in DR1, the main uncertainty is the background level. The impact is most visible in low signal-to-noise spectra. In Fig. 17, we compare the NICMOSlook extraction of three spectra of the McCarthy et al. targets with the one from DR1, all of them are low signal-to-noise spectra. The good agreement of the spectra suggests that both the relative and absolute flux calibration in DR1 is reliable.

To summarise, the accuracy of the completely unsupervised extractions of NICMOS spectra in HLA DR1 is close to what can be achieved with an interactive tool. In addition, we have taken into account contamination by sources outside the field of view of NICMOS, which is not possible using currently available interactive tools.

8.7. Comparison of Spectra with Published Data

As a quality check, we compared some extracted spectra to spectra of the same objects taken from the ground, as well as previously published extractions of spectra from the same NICMOS data. Because of strong atmospheric lines, H-band IR spectroscopy is difficult to calibrate from the ground. In particular, the region around $\lambda = 1.4 - 1.5\,\mu$ m is not accessible from the ground.

In Figs. 18 and 19, we compare spectra of two high redshift QSOs to spectra from Gemini (Stern et al., 2003) and the Italian Telescopio Nazionale Galileo (TNG, Maiolino et al., 2004). In both cases, the shape of the line and continuum around the CIII line at $\lambda \sim 1.3\,\mu$m is well reproduced. The continuum at wavelengths longer than $1.5\,\mu$m agrees to within about 5% with the Gemini data, but differs by almost a factor of two from the TNG data. We attribute this discrepancy to the uncertain calibration of the data from the ground.

Finally, in Fig. 20, we compare HLA spectra of the brown dwarf ASR 24 with the spectra extracted from the same data set by Greissl et al. (2007). ASR 24 has been observed three times with NICMOS G141 with three different roll angles. The HLA therefore contains three separate spectra of this source, namely HNG_J032911.32+311717.5_N8VM06BEQ, HNG_J032911.32+311717.5_N8VM09G5Q and HNG_J032911.32+311717.6_N8VM16S3Q. Greissl et al. (2007) did not derive a flux scale for their spectrum, this spectrum is therefore scaled to matched the HLA extraction. It can be seen that the absolute and relative flux calibration of the three HLA extractions agree to within 10%. We therefore conclude that the quality of the HLA NICMOS G141 spectra is comparable to the best extractions previously obtained from the data.

9. Summary and Conclusion

The HLA NICMOS grism project provides a database of low resolution H-band IR spectra. Most of the spectra have never been previously extracted from the HST NICMOS grism data. The database is useful for work on cool stars and emission line galaxies at z between 1.1 and 1.9, which are readily detected with the NICMOS grism. The calibration of the spectra is based on a new analysis of available calibration data and extraction of the spectra for isolated point source should be close to optimum. Confused spectra and spectra of extended objects are identified in the data release and are also of high quality. The absolute and relative flux calibration of the spectra is better than 10%, and the wavelength calibration better than 5 nm. The data release is

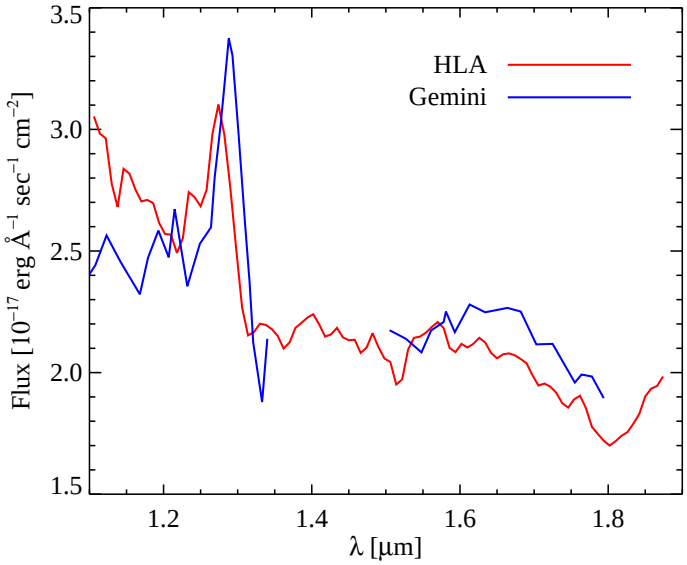

Fig. 18. Comparison of an HLA spectrum with a Gemini spectrum of the same source. The blue curve is the heavily smoothed spectrum of SDSS J083643.85+005453.3 from Stern et al. (2003), the red curve is HNG_J083643.82+005453.4_N6LE01ULQ.

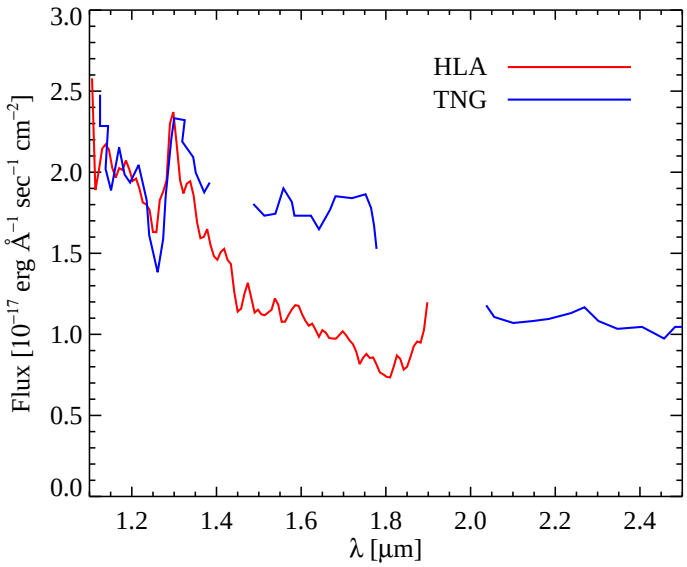

Fig. 19. Comparison of an HLA spectrum with a TNG NICS spectrum of the same source. The blue curve is the spectrum of SDSS 104433.04+012502.2 from Maiolino et al. (2004), the red curve is HNG_J104433.08-012501.6_N6LE03C7Q.

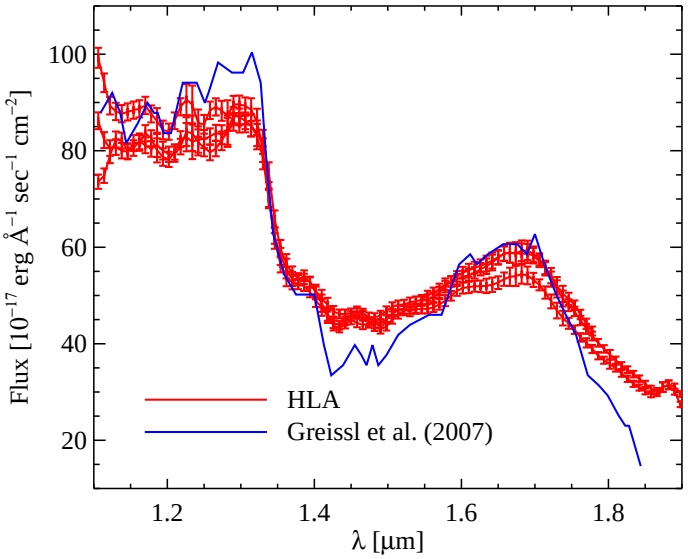

Fig. 20. Comparison of an HLA spectra of brown dwarf ASR 24 (red curves and error bars) with the spectrum published by Greissl et al. (2007) (blue curve). The latter has been scaled to match the mean flux level of the HLA data

accompanied by a wide range of auxiliary data and is available through several interfaces. We anticipate a revised data release based on a new version of the STScI CALNICA pipeline in early 2009.

Acknowledgements. This paper is based on observations made with the NASA/ESA Hubble Space Telescope, obtained from the data archive at the Space Telescope – European Coordinating Facility. We thank our HLA collaborators Brad Whitmore and the STScI and CADC HLA teams.

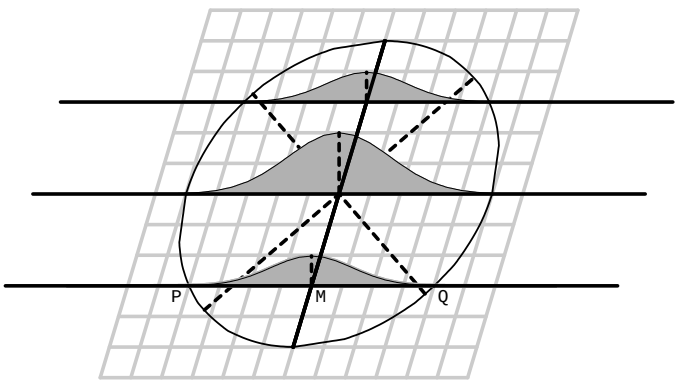

Fig. A.1. Schematic drawing of a slitless spectrum measurement.

Appendix A: Extraction of Slitless Spectra

A crucial difference between longslit and slitless spectroscopy is the selection of the light to be dispersed. For slitless spetroscopy, the light that enters the spectrograph is only limited by the object shape and the point spread function. For the extraction, we neglected the fact that the shape, as well as the point spread function, can change with wavelength. For the following discussion, we also neglect the impact of the sensitivity curves on the spectrum. A more rigorous discussion of the parameters will be presented by Lombardi et al. (2008). With the above approximations, the specific intensity of the object $I(\boldsymbol{x}, \lambda)$ at the angular position $\boldsymbol{x}$ can be expressed as the product of two functions

$$I(\boldsymbol{x}, \lambda) = I(\boldsymbol{x})F(\lambda) \ , \tag{A.1}$$

where $I(\boldsymbol{x})$, is the integrated intensity of the object as revealed by the direct image, and $F(\lambda)$ is the spectral energy distribution. $F(\lambda)$ is independent of the position $\boldsymbol{x}$. The image $\tilde{I}(\boldsymbol{x})$ generated by the slitless spectrograph is then

$$\tilde{I}(\boldsymbol{x}) = \int I(\boldsymbol{x} - \boldsymbol{r}(\lambda - \lambda_0), \lambda)\, \mathrm{d}\lambda = \int I(\boldsymbol{x} - \boldsymbol{r}(\lambda - \lambda_0))F(\lambda)\, \mathrm{d}\lambda \ , \tag{A.2}$$

where $\boldsymbol{r}$ is the dispersion vector and λ_0 a pivot wavelength. The second equality shows that the result can be written as a simple convolution, carried out along the dispersion direction, of the original image with the object spectrum.

The goal of the spectral extraction is to add all flux values of pixels that represent the same wavelength. For slitless spectroscopy, the monochromatic light from a single wavelength covers a region with the same shape as the target object on the detector. For simplicity, we assign all pixels on a straight line on the two-dimensional grism image to the same wavelength bin in the one-dimensional spectrum. The size of the objects along the dispersion direction limits the spectra resolution of slitless spectra and has the same effect as the width of the slit in a longslit spectrograph. In the following, we will refer to the box defined by the above straight line of constant wavelength and the size of the object along the dispersion direction as the "virtual slit".

For complex objects, it is unavoidable that a straight line includes pixels with flux originating from different wavelengths. The goal of the extraction procedure is to choose the direction of the virtual slit so that variations of wavelength are minimised. Our algorithm is designed to work exactly for the case of elliptical objects, i.e. in that case each wavelength bin of the one-dimensional spectrum includes only pixels with flux from a single wavelength.

As shown above in Eq. (A.2), the two-dimensional spectrum is a convolution of the object image with the one-dimensional spectrum. Spectral features will be smoothed by the object intensity profile corresponding to each line of dispersion. For example, for an object with a single emission line at λ_e so that $F(\lambda) \propto \delta(\lambda - \lambda_\mathrm{e})$, the resultant image $\tilde{I}(\boldsymbol{x})$ will be identical to $I(\boldsymbol{x})$ except for a shift along $\boldsymbol{r}$.

A natural choice for the orientation of the virtual slit is the direction of the line defined by the "centre of light" of the intensity profile along the dispersion direction. In this case, the position $\boldsymbol{x}_\mathrm{e}$ satisfies the equation

$$\int \tilde{I}(\boldsymbol{x}_\mathrm{e} + \boldsymbol{r}\ell)\ell\, \mathrm{d}\ell = 0 \ . \tag{A.3}$$

For elliptical isophotes, the image of the object can be written as a function of the elliptical radius $\rho^2 \equiv \boldsymbol{x}^\mathrm{T} A \boldsymbol{x}$,

$$I(\boldsymbol{x}) = f(\boldsymbol{x}^\mathrm{T} A \boldsymbol{x}) \ , \tag{A.4}$$

where A is a 2×2 symmetric matrix, and where, for simplicity, we have chosen the origin of our coordinates at the centre of the elliptical object. In the following, without loss of generality, we will also assume that the dispersion direction is along the horizontal axis x_1, so that $\boldsymbol{r} = (r, 0)$. For convenience, we convert to a *slanted* coordinate system $\boldsymbol{x}'$ defined by the linear transformation

$$x_1 \mapsto x_1' = x_1 + kx_2 \ , \qquad x_2 \mapsto x_2' = x_2 \ . \tag{A.5}$$

$$\boldsymbol{x} \mapsto \boldsymbol{x}' = T\boldsymbol{x} \ , \qquad T = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \ . \tag{A.6}$$

In this new coordinate system we then define the slanted direct image $I'(\boldsymbol{x}')$ as a simple remapping of $I(\boldsymbol{x})$:

$$I'(\boldsymbol{x}') = I(\boldsymbol{x}) = I(T^{-1}\boldsymbol{x}') \ . \tag{A.7}$$

Along a line parallel to the dispersion direction, the transformation T is a simple translation: in other words, for fixed x_2, x_1' is a simple shift of x_1 by kx_2.

To compute the orientation of the virtual slit, we choose the value of k in the transformation T in such a manner that the image of the galaxy in the slanted coordinate system appears as an ellipse with one of the axes oriented along the dispersion direction. This requirement is equivalent to choosing the matrix $T^{-\mathrm{T}}AT^{-1}$ to be diagonal, which is guaranteed if $k = A_{12}/A_{11}$. With this choice, we find

$$T^{-\mathrm{T}}AT^{-1} = \begin{pmatrix} A_{11} & 0 \\ 0 & \det(A)/A_{11} \end{pmatrix} \ . \tag{A.8}$$

In this new coordinate system the virtual slits are perpendicularly to the dispersion direction, because the object is now symmetric along the axis x_2'. The corresponding lines in the original coordinate system can be obtained by transforming back the vertical lines, and as graphically illustrated in Fig. A.1, these lines can be obtained by joining the two tangent points of the ellipse describing the direct object with lines parallel to the dispersion direction.

Note that the direction of the virtual slits is never along one of the axes of the elliptical isophotes except in the trivial case where the object is already oriented along the dispersion direction ($A_{12} = k = 0$). The angle with respect to the vertical formed by the lines of constant wavelength is given by $\tan\beta = k = A_{12}/A_{11}$, and thus for highly elongated objects it can approach $\pm\pi/2$.

References

Adelman-McCarthy, J. K., et al. 2007, ApJS, 172, 634

Bertin, E. & Arnouts, S. 1996, A& A, 117, 393

Bohlin, R. 2003, 2002 HST Calibration Workshop, ed. S. Arribas, A. Koekemoer, & B. Whitmore, (Baltimore:STScI), 115

Bohlin, R. C. 2007, in The Future of Photometric, Spectrophotometric, and Polarimetric Standardization, ASP Conf. Series, Vol. 364, p. 315 ed. C. Sterken; astro-Ph 0608715

Bohlin, R.C., Riess, A. & de Jong, R. 2006, Instrument Science Report, NICMOS, 2006-002

Bohlin, R.C., Lindler, D. & Riess, A. 2005, Instrument Science Report, NICMOS, 2005-002

Bushouse, H. 1997, The 1997 HST Calibration Workshop with a New Generation of Instruments, 223, 223

Dahlen, T., et al. 2008, Instrument Science Report, NICMOS, 2008-002

Freudling, W. 1997, Space Telescope European Coordinating Facility Newsletter, Volume 24, 7, 24, 7

Fruchter, A. S., Hook, R. N. 2002, PASP, 114, 144

Greissl, J., Meyer, M. R., Wilking, B. A., Fanetti, T., Schneider, G., Greene, T. P., & Young, E. 2007, AJ, 133, 1321

Horne, K. 1986, Proc.Astr.Soc.Pacific, 98, 609

Koekemoer, A.M., Fruchter, A.S., Hook, R.N. & Hack,W. 2002, Proceedings of the 2002 HST Calibration Workshop, STScI, 337

Koekemoer, A.M., Fruchter, A.S., Hook, R.N., Hack W., & Hanley, C. 2006, The 2005 Calibration Workshop, eds. A. Koekemoer, P. Goudfrooij, & L. Dressel, L., 423

Kümmel, M., Walsh, J.R., Pirzkal, N., Kuntschner, H., & Pasquali, A., 2008, PASP, submitted

Kümmel, M., Kuntschner, H., & Walsh, J. R. 2007 , Space Telescope – European Coordinating Facility Newsletter, 43, 8

Lasker, B. M., Sturch, C. R., McLean, B. J., Russell, J. L., Jenkner, H., & Shara, M. M. 1990, AJ, 99, 2019

Lombardi et al., 2008, in preparation

MacKenty, J. 1997a, HST Proposal, 7907

MacKenty, J. 1997b, HST Proposal, 8082

Malkan, M. 2003, HST Proposal, 9865

Malkan, M. 2004, HST Proposal, 10226

Maiolino, R., Oliva, E., Ghinassi, F., Pedani, M., Mannucci, F., Mujica, R., & Juarez, Y. 2004, A& A, 420, 889

McCarthy, P. J., et al. 1999, ApJ, 520, 548

McDowell, J. & Tody, D. (eds.) 2007, IVOA Spectral Data Model Version 1.01, http://www.ivoa.net/Documents/PR/DM/SpectrumDM-20070515.html

McLean, B. J., Greene, G. R., Lattanzi, M. G., & Pirenne, B. 2000, Astronomical Data Analysis Software and Systems IX, 216, 145

Mobasher, B. & Roye, E. eds. 2004, HST Data Handbook for NICMOS V.6

Monet, D. G., et al. 2003, AJ, 125, 984

Pirzkal, N., Xu, C., Malhotra, S., et al. 2004, ApJS, 154 501

Schultz, A. B., et al. 2003, Proc. SPIE, 4850, 858

Stern, D., Hall, P. B., Barrientos, L. F., Bunker, A. J., Elston, R., Ledlow, M. J., Raines, S. N., & Willis, J. 2003, ApJL, 596, L39

Storrs, R., Hook, R., Stiavelli, M., Hanley, C., & Freudling, W., 1999, Instrument Science Report, NICMOS, 99-005

Stoehr, F., et al. 2007, Space Telescope European Coordinating Facility Newsletter, 42, 4

Thompson, R. I., & Freudling, W. 2002, HST Calibration Workshop: Hubble after the Installation of the ACS and the NICMOS Cooling System, 240

Skrutskie, M. F., et al. 2006, AJ, 131, 1163

Xu, C., & Mobasher, B. 2003, Instrument Science Report, NICMOS, 2003-009

Zacharias, N., Urban, S. E., Zacharias, M. I., Wycoff, G. L., Hall, D. M., Monet, D. G., & Rafferty, T. J. 2004, AJ, 127, 3043