Instrument Science Report WFC3 2023-07
Instrument Science Report ACS 2023-05

HSTaXe - ACS & WFC3 Cookbook Tutorials

B. Kuhn, D. Som, A. Pidgeon,
N. Hathi, R. Ryan, R. Avila, N. Pirzkal

September 19, 2023

Abstract

This report describes a collection of six Jupyter Notebooks, released on the **HSTaXe** *GitHub repository in Spring 2023, that demonstrate data reduction using STScI's official slitless spectroscopy software,* **HSTaXe**. *These 'cookbooks' present examples of how to preprocess data from ACS and WFC3 slitless-spectroscopic modes and use the core* **HSTaXe** *routines to extract 1D spectra. The specific preprocessing procedures explained here and in the cookbooks are meant to highlight three steps of the data analysis process users should consider to obtain optimal spectral extraction with* **HSTaXe**. *The three steps include a custom multi-component background subtraction for WFC3/IR grism data, embedding subarray data into a full-chip image, and checking that the active World Coordinate System (WCS) of dispersed images matches the corresponding direct images. In addition to these preprocessing steps, we also address installation methods, the general cookbook workflow, advanced fluxcube extraction, and* **HSTaXe** *output files.*

1 Introduction

The aXe software (Pirzkal, Pasquali, and Demleitner 2001) was developed in 2000 as a stand-alone C-based modular environment to extract spectra from the Advanced Camera for Surveys (ACS) grism and prism spectroscopic modes. It was originally designed to be instrument agnostic and supported ACS/Solar Blind Channel (SBC), ACS/High Resolution Channel (HRC), and ACS/Wide Field Channel (WFC) observations. In 2003, new features such as quantitative contamination were added. As these relied on IRAF/PyRAF routines, aXe was distributed as part of the Space Telescope Science Data Analysis System (STSDAS) software package (Kümmel, Walsh, et al. 2009). Since aXe was developed to be instrument-independent, with all instrument-specific parameters defined in a single configuration file, it has been successfully adapted for processing/analyzing Near-Infrared Camera and Multi-Object Spectrometer (NICMOS), and Wide Field Camera 3 (WFC3) slitless data. As part of Hubble Legacy Archive (HLA) projects, aXe was crucial for spectral extraction in NICMOS G141 (Freudling et al. 2008) and ACS/WFC G800L (Kümmel, Kuntschner, et al. 2009). In fact, aXe has been an important tool for several large science programs, such as GRAPES (ACS/WFC; Pirzkal, Xu, et al. 2004), PEARS (ACS/WFC; Pirzkal, Burgasser, et al. 2009), and WISPS (WFC3/IR; Atek et al. 2010).

Gradually, STScI deprecated support for IRAF/PyRAF (Ogaz and Tollerud 2018), thereby imposing impediments on the utilization of the aXe software, which was STScI's official software for reducing slitless grism/prism data from the ACS and WFC3 instruments. Under the leadership of Megan Sosey and Nor Pirzkal at STScI, an updated version of aXe, called HSTaXe, was developed in Python/C and is completely independent of IRAF/PyRAF. The first official release of HSTaXe v1.0 was in February 2021, which primarily focused on full functionality for the most frequently used instruments, ACS/WFC and WFC3/IR.

Since the first release, extensive work has been done to make HSTaXe compatible with ACS/SBC prism data and WFC3/UVIS G280 grism data. An updated version of HSTaXe v1.0.5[1] (Sosey et al. 2023) was released in July 2023 and now includes full functionality for all slitless observing modes of the ACS and WFC3 instruments. To support the updated HSTaXe software, we have also released example "cookbook" tutorials (Python Jupyter Notebooks[2]; Kluyver et al. 2016) for all ACS and WFC3 slitless observing modes. These cookbooks contain new tools/scripts for preprocessing grism and prism data, (applied before calling the HSTaXe software), which cover substantially improving WFC3/IR background subtraction, embedding subarray images so they are compatible with HSTaXe, and correcting for World Coordinate System (WCS) mismatch between the direct and dispersed image.

This ISR discusses these new cookbooks and preprocessing steps in detail. We encourage the reader to employ the most recent version of the HSTaXe software (v1.0.5 at time of publication), and use the provided cookbooks and software tools when reducing ACS and WFC3 slitless data. We will continue to maintain the cookbooks to ensure their compatibility with any future changes to the HSTaXe software.

2 Installation

The HSTaXe software is hosted in an STScI GitHub repository[3] and distributed through the Python Package Index (PyPI[4]). One of the most important things to know is that **the HSTaXe software is not compatible with stenv**[5], the official environment, provided and maintained by STScI, for calibrating/analyzing both HST and JWST data. Instead, users

[1] HSTaXe Zendo release page: https://zenodo.org/record/8136948
[2] Jupyter homepage: https://jupyter.org/
[3] HSTaXe GitHub repository: https://github.com/spacetelescope/hstaxe
[4] HSTaXe PyPI webpage: https://pypi.org/project/hstaxe/
[5] stenv GitHub repository: https://github.com/spacetelescope/stenv

will need to go to the HSTaXe GitHub repository and follow one of two installation methods that will create a new, unique environment. Both installation methods require a preinstalled Python distribution and package manager; we recommend Anaconda[6].

1. **Recommended:** Install HSTaXe via the cookbook_env.yml file[7], located in the "cookbooks" folder. This method also installs all other Python packages needed for running the cookbooks (e.g. crds, jupyter, and wfc3tools[8]).

2. Install only the packages needed to run the HSTaXe software itself. This method is detailed in the "Quickstart" section of the README.md file[9], as well as the "Installation" section in the official documentation[10].

Users should install HSTaXe in a new virtual conda environment to avoid version conflicts with other packages in a pre-existing environment. The two installation methods enumerated above will create a new conda environment and install HSTaXe into it. At current, HSTaXe is unfortunately **not compatible with the Windows operating system**.

One reason we recommend installing HSTaXe with the cookbook_env.yml file (method number 1 above) is that installation is validated directly by the cookbooks themselves. Every cookbook has its own example data that is accessible via the STScI Box application[11]. Additionally, each cookbook has an accompanying HTML version that shows the expected output for every code cell. Once the data, configuration files, catalogs, etc. are downloaded, the cookbooks are designed to run from start to finish without any user input. Users can validate the HSTaXe installation by running any given cookbook, checking that everything runs successfully, and comparing the outputted 2D image stamps (STP files) and 1D extracted spectra (SPC files) to those in the HTML cookbook file (see Section 5 for information about output files).

While the cookbooks are designed to run from start to finish without any user input, *it is crucial that you take time to read the information in each cookbook carefully, especially if you are a new HSTaXe user.* Any issues with the installation or running of the cookbooks should be reported to the HST General Help Desk[12].

3 General Cookbook Workflow

The HSTaXe cookbooks are a set of Jupyter Notebook tutorials that demonstrate the preprocessing and spectral extraction process for data from all ACS and WFC3 slitless observing modes. These example notebooks are hosted on the HSTaXe GitHub repository under the cookbooks folder[13]. Within the cookbook folder, the notebooks are separated by instrument

[6]Anaconda homepage: https://www.anaconda.com/
[7]Cookbook yaml file: https://github.com/spacetelescope/hstaxe/blob/main/cookbooks/cookbook_env.yml
[8]wfc3tool GitHub repository: https://github.com/spacetelescope/wfc3tools
[9]HSTaXe README file: https://github.com/spacetelescope/hstaxe#readme
[10]Official HSTaXe documentation: https://hstaxe.readthedocs.io/en/latest/hstaxe/installing.html
[11]STScI's official Box page: https://stsci.app.box.com/
[12]HST Help Desk: https://stsci.service-now.com/hst
[13]HSTaXe cookbook folder: https://github.com/spacetelescope/hstaxe/tree/main/cookbooks

with one folder for ACS and another for WFC3. Inside the ACS folder, there are (currently) three different cookbooks that cover the following topics:

1. Spectral extraction with full-frame data from ACS's Solar Blind Channel ,
2. Spectral extraction with full-frame data from ACS's Wide Field Channel (WFC) , and
3. Spectral extraction with subarray data from ACS/WFC

The cookbooks in the WFC3 folder are separated in a very similar way. Within the folder, there are three cookbooks that encompass the following topics:

1. Spectral extraction with data from WFC3's IR channel (with an included advanced extraction example),
2. Spectral extraction with full-frame data from WFC3's UVIS channel , and
3. Spectral extraction with subarray data from WFC3/UVIS

Each cookbook has its own example data, hosted on Box and linked within the given notebook. Once the example data are downloaded to the correct locations, the cookbooks demonstrate how the directories can be organized and create copies of the files. It is always recommended to create copies of the original files because `HSTaXe` will edit them as different routines are called. After data download and organization, the next preprocessing steps presented in the cookbooks are determined by the type of data being analyzed. In Figure 1, we present a flowchart to help illustrate the steps based on the input data. The flowchart highlights (in green) the three major preprocessing steps, discussed in detail here, that one may need to undertake before starting the basic or advanced `HSTaXe` workflow. These steps include:

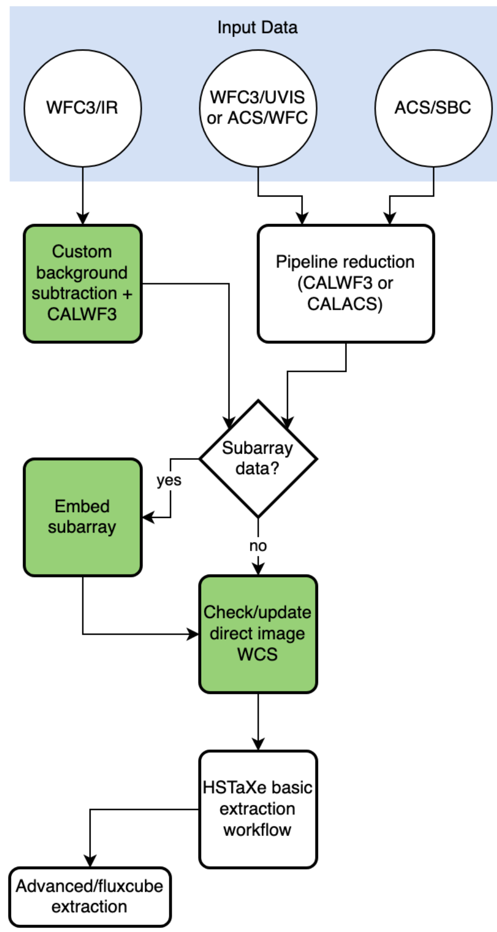

Figure 1: Flowchart illustrating the recommended preprocessing steps one should take to achieve optimally extracted spectra with `HSTaXe`.

1. a custom multi-component background subtraction for WFC3/IR data (Section 4.1),
2. an embedding procedure for any subarray data (Section 4.2), and

3. a function to check (and update if needed) the WCS of the dispersed and direct images (Section 4.3)

After completing the necessary preprocessing steps, the data are ready for the core HSTaXe workflow. These steps are not detailed in the flowchart, but they are presented and explained in each cookbook. Here, we list each of the core HSTaXe workflow steps to achieve basic extraction.

1. Use AstroDrizzle (Hack et al. 2023) to create combined images of the grism and direct exposures for improved source detection accuracy.
2. Generate a catalog with SExtractor (Bertin and Arnouts 1996) using the aXe-SExtractor configuration files.
3. Create input object lists with the iolprep routine within axetasks.
4. Call the axeprep routine within axetasks to prepare the exposures for spectral extraction.
5. Run the axecore routine within axetasks to perform the spectral extraction.

The motivation behind each of these steps is detailed in the cookbooks, so users should carefully read all text between the code cells. For example, users with full-frame ACS/WFC or WFC3/UVIS data should be mindful of the chip-dependent configuration and input object list files necessary for running HSTaXe.

Within each of the above steps, there are many optional and mandatory parameters to pass to each routine. The official documentation lists all parameters for axeprep (Section 6.3) and axecore (Section 6.4). Background subtraction is an example of an optional parameter for both routines. While HSTaXe is capable of performing a global (via axeprep) and/or local (via axecore) background subtraction, it is not recommended when analyzing WFC3/IR data. Instead, users should turn off global background subtraction, which is turned on by default in the axeprep routine, and run WFC3 Backsub[14] (a Python script) on the RAW, uncalibrated IR grism FITS files to remove the dispersed multi-component background (see Section 4.1 below for more details). When analyzing ACS/WFC data, users are limited to only global background subtraction as local subtraction is not fully functional. Finally, users analyzing WFC3/UVIS data are only able to perform local background subtraction because there is no G280 master sky background image available, although, at the time of publication, one is being developed (Pagul et al. 2023 in review).

Beyond these basic HSTaXe workflow steps and parameters, there is also an optional advanced extraction method. This approach involves creating multiple mosaics and generating fluxcubes to more accurately estimate contamination and improve the signal-to-noise of the extraction. This procedure is described in Section 6 and demonstrated in Section 4.2 of the WFC3/IR extraction cookbook[15].

[14]WFC3 Backsub Python script: https://stsci.box.com/s/xrz16v1tv58y8vcfy8zlrn83jrpnam19
[15]WFC3/IR cookbook: https://github.com/spacetelescope/hstaxe/tree/main/cookbooks/WFC3/IR_extraction

4 Recommended Preprocessing Steps

In the following subsections, we describe preprocessing steps (shown by green cells in Figure 1) that users should consider before using `HSTaXe`. While these steps are not actually part of the `HSTaXe` software, we include them in our cookbooks for the benefit of the HST user community, to enable optimal extraction of slitless spectroscopic data from HST.

4.1 Background Subtraction for WFC3/IR data

For both G102 and G141, the background sky signal is variable and also made of multiple components. Currently, the WFC3 calibration pipeline, `calwf3`, does not have the capability to model and remove the dispersed 2D background. Therefore, **we highly advise that observers use the WFC3 Backsub Python script to process uncalibrated IR grism images (RAWs) into calibrated, flat-fielded files (FLTs)**. WFC3 Backsub is explicitly designed to assess the level of each of the three background components (zodiacal, 1.083 μm HeI emission, and scattered light) and remove the contributed signal. WFC3 Backsub still relies on and uses `calwf3` for bias correction and dark subtraction, but it employs custom reference files, (not available in the Calibration Reference Data System (CRDS)), to measure and remove the multiple sky components before the final "up-the-ramp" fitting occurs in `calwf3`. A description of the three background components and the methods used in WFC3 Backsub can be found in WFC3 ISR 2020-04 (Pirzkal and Ryan 2020).

In Figure 2 we show a G141 grism image calibrated without any background subtraction (left) and the same image calibrated with WFC3 Backsub (right). While the effects of a dispersed background can be seen in the single exposure FLT files, Figure 2 shows drizzle products that combine four G141 images, which amplify the effects. There is a clear and dramatic difference in the background features and smoothness. For example, the artifact on the lower right side of the image on the left, known as the "wagon wheel" (Bushouse 2008), is almost completely removed from the corresponding location in the image on the right when using WFC3 Backsub. The vertical banding seen on the left and right edges of the image without background subtraction is characteristic of a dispersed multi-component background. The image that uses WFC3 Backsub, has very few signs of the vertical banding left over. We note that here, as should be done when drizzling WFC3/IR grism images, `AstroDrizzle`'s background subtraction process was turned off by setting the `skysub` argument to `False`.

WFC3/IR grism users interested in using WFC3 Backsub should refer to the IR extraction cookbook for more details. Section 3.1 of the cookbook walks through how to set up and use WFC3 Backsub.

4.2 Embedding Procedure for Subarray data

While both ACS and WFC3 instruments support observing with default and custom subarrays, the `HSTaXe` software requires an additional step before spectral extraction. This step involves embedding the subarray image into a full *chip* image (e.g. $\sim$ 2K$\times$4K pixels for WFC3/UVIS and ACS/WFC). Once the direct and grism FLT files are embedded into a full chip image the rest of the preprocessing and spectral extraction with `HSTaXe` can occur.

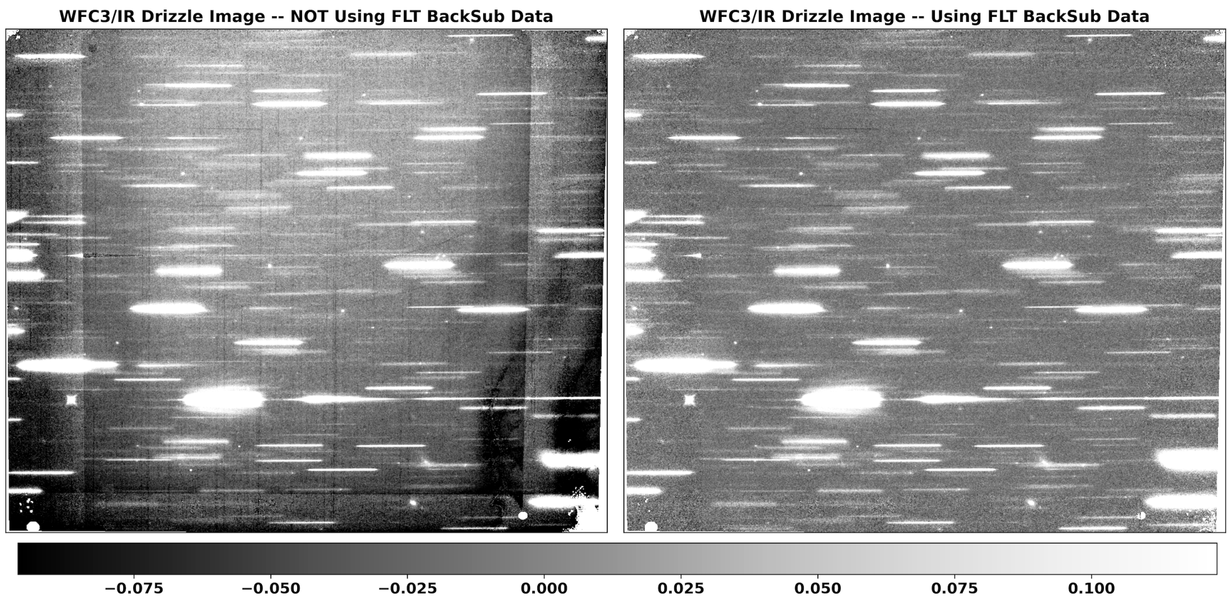

Figure 2: WFC3/IR G141 drizzled images that were generated without (left) and with (right) data processed by the custom WFC3 Backsub Python script . The same four individual exposures went into creating each drizzle product above, except the calibrated FLT files that created the image on the right were corrected for the dispersed, multi-component sky background using WFC3 Backsub. The units of the color bar at the bottom are in e^-/s and the total exposure time is 4211.7 s. Data come from Visit 23 of GO/DD 11359; PI: Robert O'Connell (`ib6o23*.fits`).

Figure 3 illustrates an example UVIS (left) and WFC (right) subarray before (top) and after (bottom) the embedding process. The images on top are the original subarray FLTs with a size of 768×4096 pixels (left) and 2048×2048 pixels (right). After running the embedding routines we are returned a larger full chip image seen at the bottom left and right of Figure 3, which have the sizes of one UVIS and WFC charge-coupled device (CCD), 2051×4096 and 2048×4096 pixels respectively.

For WFC3 data (UVIS or IR) the embedding process can be completed using the STScI-developed Python script `embedsub`, which is hosted on the `wfc3tools` GitHub repository. Users with ACS subarray data should use a similar, but separate, script called `embed_subarray_acs.py`, which is hosted on the STScI Box application[16]. Both embedding scripts require users to provide the FLT image as well as the corresponding telemetry support (SPT) file. As detailed in Section 3, the ACS and WFC3 teams have provided instrument-specific cookbooks for handling subarray data.

4.3 Checking and Correcting for WCS Mismatch

The WCS for an image describes the mapping between its pixel (image) coordinates and physical coordinates (e.g., RA and Dec). Since late 2019, MAST data products for ACS and

[16]ACS subarray embedding Python script: https://stsci.box.com/s/hprbymg0g0cvacl559fqgju6o0gbfdzp

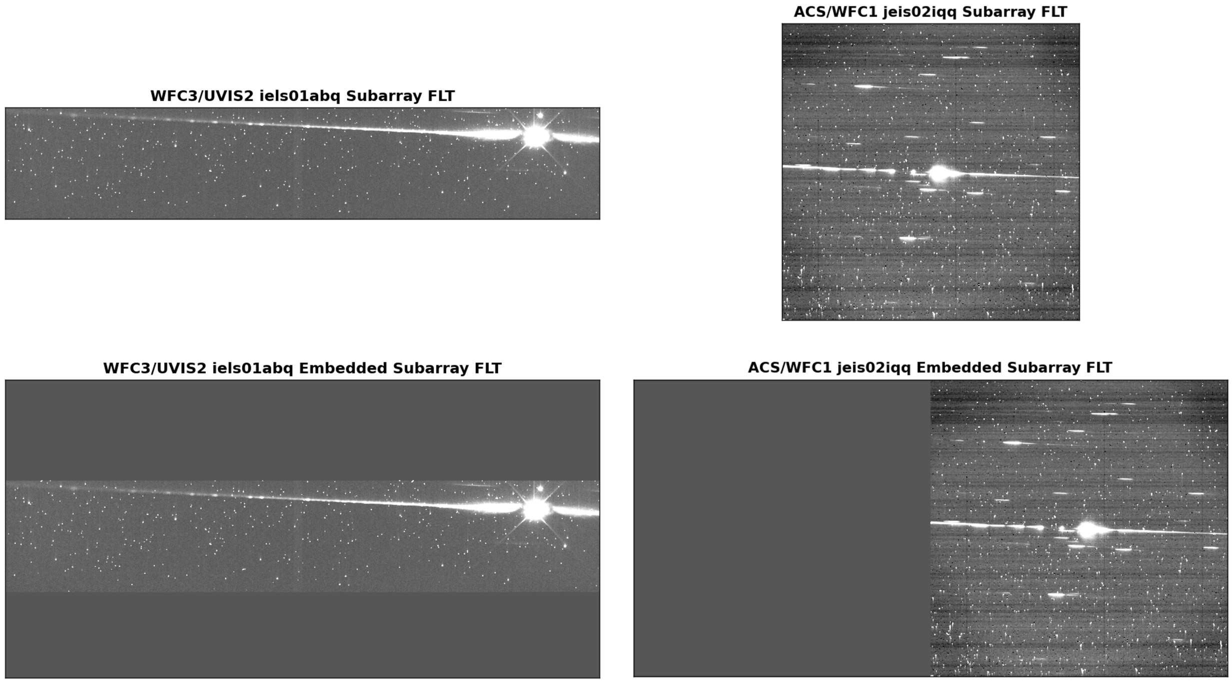

Figure 3: Examples of a WFC3/UVIS G280 (left) and ACS/WFC G800L (right) subarray FLT images before (top) and after (bottom) the embedding process. The original UVIS and WFC subarray images on the top row have a shape of 768×4096 and 2048×2048 pixels respectively. After embedding the subarray into a full chip image the final shape is 2051×4096 pixels for UVIS and 2048×4096 pixels for WFC. The embedding step is crucial as `HSTaXe` does not support subarray data (ACS or WFC3) that is not first embedded into a full chip image. The UVIS data is from Visit 01 of CAL/WFC3 16581; PI: Debopam Som and the WFC data is from Visit 02 of CAL/ACS 16474; PI: Dean Hines.

WFC3 include improved absolute astrometry provided via the WCS in the image headers (Mack et al. 2022). As part of these improvements, *all* images now include in their WCS an "a priori" correction - an update of the guide star coordinates to match either the most updated HST Guide Star Catalog (GSC v2.40) or the Hubble Source Catalog (HSC v3.0). An additional "a posteriori" correction has been applied whenever alignments between detected sources in the image and a select set of reference catalogs (e.g. Gaia eDR3) were possible. However, the WCS contains *only* the "a priori" correction for observations obtained using modes not permitting such alignments (e.g., slitless spectroscopy and moving target observations), or for which the alignment failed due to a lack of sources in either the image or the reference catalog.

It is customary for slitless spectroscopic observations to obtain direct images to serve as astrometric standards for the slitless spectroscopic exposures; this linkage is achieved using the WCS information associated with the images. Each spectral extraction workflow described in this work defines spectral traces and sets wavelength zero-point for the spectroscopic images using a source catalog generated from distortion-corrected direct images, and in doing so, assumes the same WCS across all images. Therefore, to achieve optimal spectral extraction, it is imperative to first ensure that the WCS in the input direct and dispersed images match.

We note the possibility of ACS and WFC3 data retrieved from MAST having a WCS mismatch between spectroscopic and direct images. The WCS of the spectroscopic images, always containing only the 'a priori' corrections, may be offset from their direct image counterparts if the latter are aligned to an external reference catalog. Such misaligned data could lead to erroneous spectral extraction, especially affecting the accuracy of the wavelength calibration. Figure 4 shows an example of such inaccurate wavelength calibration extracted from WFC3 data with mismatched WCS. From Gaussian fits to the emission features, the spectrum impacted by WCS misalignment, in this example, was estimated to be blue-shifted by ~ 20 $\AA$ in comparison with the WCS-matched extraction.

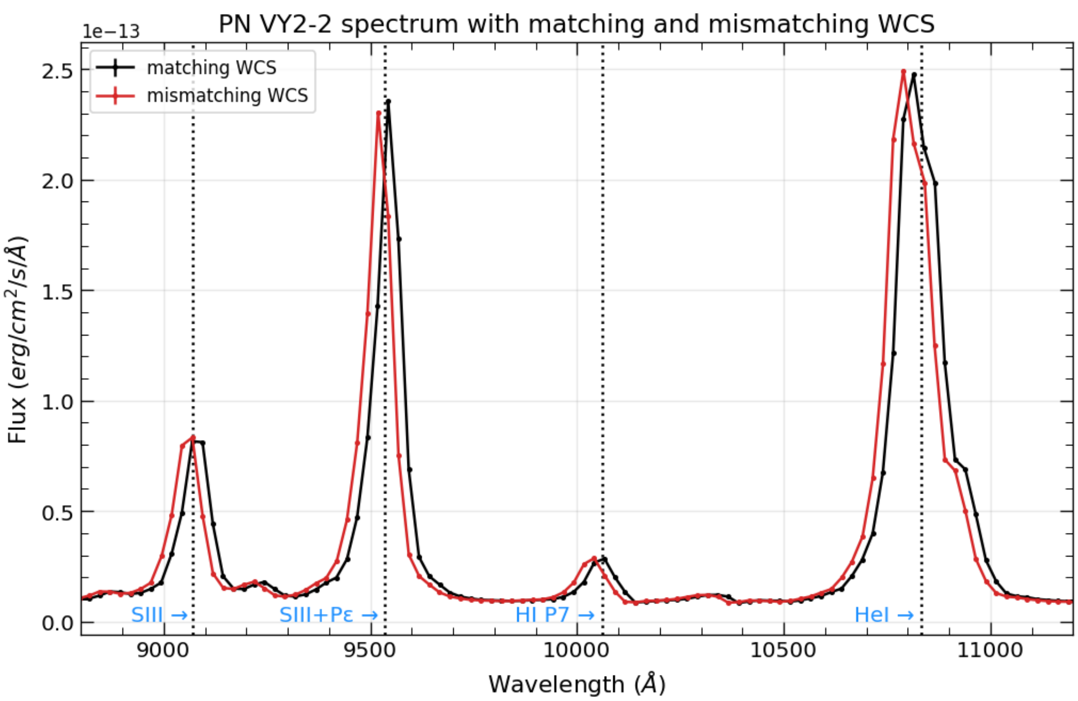

Figure 4: Impact of WCS misalignment between direct and spectroscopic images on the wavelength calibration of spectral extraction using `HSTaXe`. The spectrum in red represents WFC3/IR G102 data of planetary nebula VY2-2 (from WFC3 Calibration Program 16582; PI: Debopam Som) extracted via `HSTaXe` using direct and spectroscopic images with 'broken' WCS alignment. The spectrum in black represents the same data extracted using `HSTaXe` but after the WCS mismatch has been corrected in preprocessing (after MAST retrieval and before extraction). The vertical dotted lines represent the rest wavelengths (i.e., expected locations) of known emission features (labeled on the bottom of the plot in blue) from Table 2 of Bohlin, Deustua, and Pirzkal 2015.

We, therefore, **strongly recommend that users check WCS consistency between direct and dispersed images retrieved from MAST;** any mismatch should be corrected prior to attempting spectral extraction. Examples of how to perform this preprocessing step, using a custom-built function, have been included in all HSTaXe cookbooks discussed in this report. This function checks for WCS consistency across all input images and, when a disagreement is detected, calls `updatewcs`[17] to roll back all the solutions to the original distortion-corrected WCS.

5 Overview of HSTaXe Outputs from Basic Extraction

After successful completion of the necessary preprocessing steps (Section 4) and basic HSTaXe workflow (Section 3), one should expect to see the `OUTPUT` directory populated with several files corresponding to each input spectroscopic exposure. Table 1 lists these file categories along with a short description. Users of the cookbooks may find this information useful in identifying and reviewing the subsequently generated data products, but we also recommend reviewing Section 6 of the official documentation, which provides a complete list of files generated by the various HSTaXe tasks.

Two types of files are of particular interest in examining spectra identified and extracted by HSTaXe; the one-dimensional extracted spectra (SPC) file and the two-dimensional stamp image of extracted traces (STP) file. An SPC file, generated for each input spectroscopic image file, contains 1D spectral information for all dispersion orders extracted for all sources provided in the input object catalog. The corresponding STP file contains 2D spectral stamp images for the extracted orders present in the SPC file. The bottom panel of Figure 5 shows 1D spectral extractions for the same object and dispersion order but from three separate SPC files corresponding to three input spectroscopic images. The top panel of the figure shows the corresponding 2D spectral stamp images. Please see Section 8 of the HSTaXe documentation for more information about the structures of various files required by, interacted with, and generated by HSTaXe.

Table 1: HSTaXe Output Files from Basic Extraction

File Suffix	Description
cat	Input object catalog
OAF	Original aperture file
BAF*	Background aperture file
BCK.fits*	Background image
BCK.PET.fits*	Pixel extraction tables for the background
CONT.fits	Contamination estimate for spectra
PET.fits	Pixel extraction tables
SPC.fits	1D extracted spectra
STP.fits	2D stamp image of extracted traces
opt.SPC.fits†	Optimally extracted 1D spectra
opt.WHT.fits†	Optimal weight image

* Files produced if the argument `back=True` in the `axecore` task.

† Files produced if the argument `weights=True` in the `axecore` task.

[17] Official documentation for `updatewcs`: https://stwcs.readthedocs.io/en/latest/updatewcs.html

The ACS/WFC and WFC3/UVIS detectors are both comprised of two physically separate ∼2K×4K-pixel CCDs, arranged such that one CCD is located above another. These CCDs are named according to their physical position from top to bottom: the constituent CCDs of ACS/WFC are named WFC1 and WFC2, while WFC3/UVIS has UVIS1 and UVIS2.

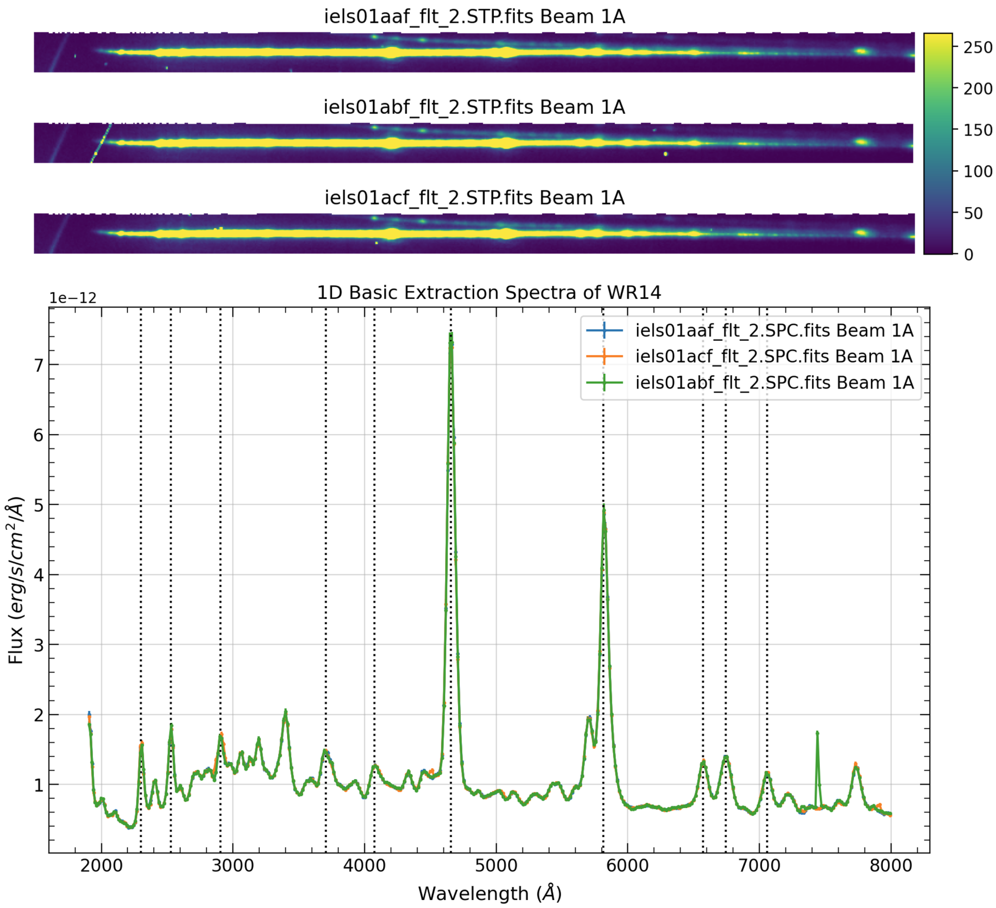

Figure 5: Examples of the contents of STP and SPC files. The top panel shows 2D spectral stamp images of the $+1^{st}$ WFC3/UVIS G280 dispersion order from the Wolf Rayet star WR-14, retrieved from three STP files corresponding to three different input spectroscopic images: `iels01aaf_flt.fits`, `iels01abf_flt.fits`, and `iels01acf_flt.fits`; from WFC3 Calibration Program 16581; PI: Debopam Som. The bottom panel shows the corresponding extracted 1D spectra, retrieved from the SPC files.

For a given full-frame observation, each CCD has its own science, error, and data quality flag extensions within the FLT FITS file. Counterintuitively, the second chip's data is found in the first set of extensions, and the first chip's data is found in the second set of extensions. As an example, if indexing the extensions of a WFC3/UVIS full-frame image, which has (a minimum of) seven extensions, the science data for UVIS2 will be in the first extension while the corresponding data for UVIS1 is found in the fourth extension. Every extension also has a descriptive name (i.e. "PRIMARY", "SCI", "ERR", and "DQ"). Since there are two sets of "SCI", "ERR", and "DQ" extensions, UVIS1 science data can also be found by accessing the “SCI,2” extension, while accessing the “SCI,1” extension leads to UVIS2 science data. “ERR” and “DQ” extensions follow the same pattern just described for the “SCI” arrays.

When processing either ACS/WFC or WFC3/UVIS full-frame data, `HSTaXe` requires two input object catalogs that follow a specific naming convention: `*_flt_1.cat` and `*_flt_2.cat`. In this case, the input catalog file numbering follows the FITS file structure and not the physical location/name of the CCDs, i.e. `*_flt_1.cat` contains data from the second/bottom CCD (WFC2 or UVIS2).

`HSTaXe` also produces output files for full-frame ACS/WFC and WFC3/UVIS data that are dependent on the particular CCD; notably, they are named depending on the FITS file index at which the science extension occurs (incremented by one), not the name of the CCD itself or the order of the science extensions (“SCI,1” and “SCI,2”). `HSTaXe` output files follow a naming pattern wherein CCD2 data is stored in files with the suffix “_2”, and CCD1 data is stored in files with the suffix “_5”. For example, when processing the full-frame WFC3/UVIS spectroscopic observation with the rootname `ibr502ibq`, `HSTaXe` will produce two basic extraction 1D spectra files. Data from UVIS2 will be stored in `ibr502ibq_flc_2.SPC.fits`, and data from UVIS1 will be stored in `ibr502ibq_flc_5.SPC.fits`. As seen in Figure 5, the fact that the file names contain “_2” indicates that they are UVIS2 data.

6 Advanced Extraction Procedure for WFC3/IR

As an addition to the basic extraction process demonstrated in the WFC3/IR cookbook, we also provide an example of a more advanced technique that provides a better estimate of contamination of the extracted spectra by using a “fluxcube” contamination model. The fluxcube is a multi-extension FITS file that contains direct images in multiple filters as well as a segmentation map; the segmentation map is itself an image wherein each pixel has an integer value corresponding to the unique detected object to which it belongs. This model is used by `HSTaXe` to generate a simulated grism image. For each detected object in this image, a model spectral contribution is generated and subtracted from it. Spectra are then extracted from the simulated image alongside the real grism images. Because the spectra for each source in the simulated image have had the contributions from the simulated objects themselves subtracted out, the information left behind contains the contribution of all *other* sources in the field to the extracted spectra. This method generates a quantitative estimate of the contamination of each source spectrum, which can be viewed as a 2D stamp and extracted into a 1D array, as shown in Figure 6.

During basic spectral extraction (as shown in the IR cookbook), 1D spectra are extracted from the 2D spectral stamps (see Section 5) directly after the stamps themselves are extracted from the grism images. In this advanced extraction procedure, we can first use DrizzlePac, implemented within `HSTaXe` as the `axedrizzle` task, to combine stamps from multiple grism images. As part of the drizzling process, stamp images are all mapped onto a uniform wavelength scale with dispersion along the x-axis and cross-dispersion along the y-axis. Cosmic rays and outlier pixels are also rejected, and the stamps' pixels are weighted to account for different exposure times across the individual grism images. Signal-to-noise is also improved by combining multiple images. The 1D spectra extracted from these drizzled stamps can be much cleaner than those produced by the basic extraction method, but it also

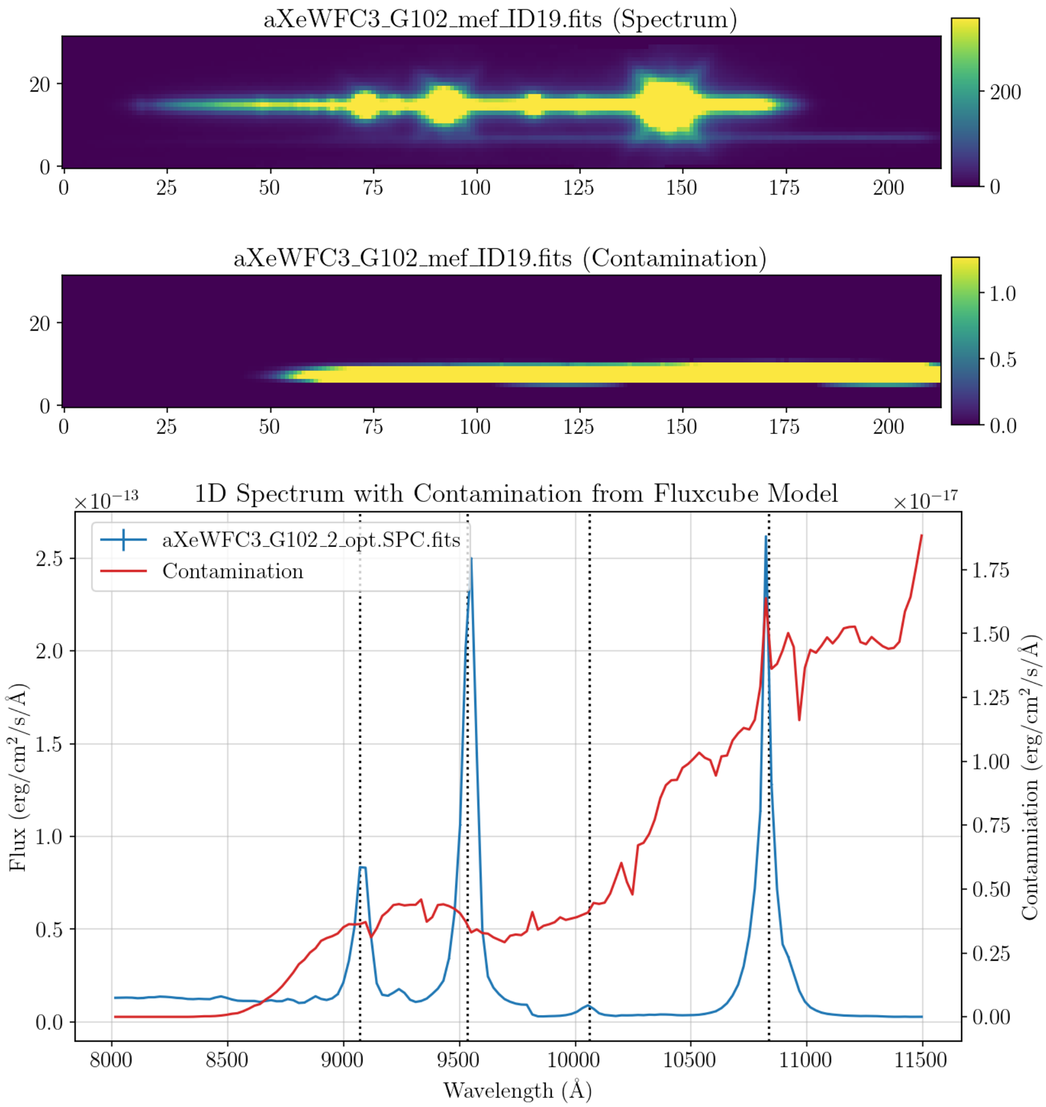

Figure 6: Contamination estimates produced by `HSTaXe` using the fluxcube model. The top panel shows 2D image stamps for the source spectrum, with a contaminating spectrum visible within the frame; the middle panel shows the contamination stamp; and the bottom panel shows the 1D extracted spectrum and contamination. This data comes from WFC3 Calibration Program 16582; PI: Debopam Som.

has the potential to adversely impact the results if the alignment of the stamps is not done properly. Results using drizzled images should therefore be carefully compared to those from the basic method. An example is shown in Figure 7; in this case, drizzling produced a better quality spectrum than the constituent FLTs.

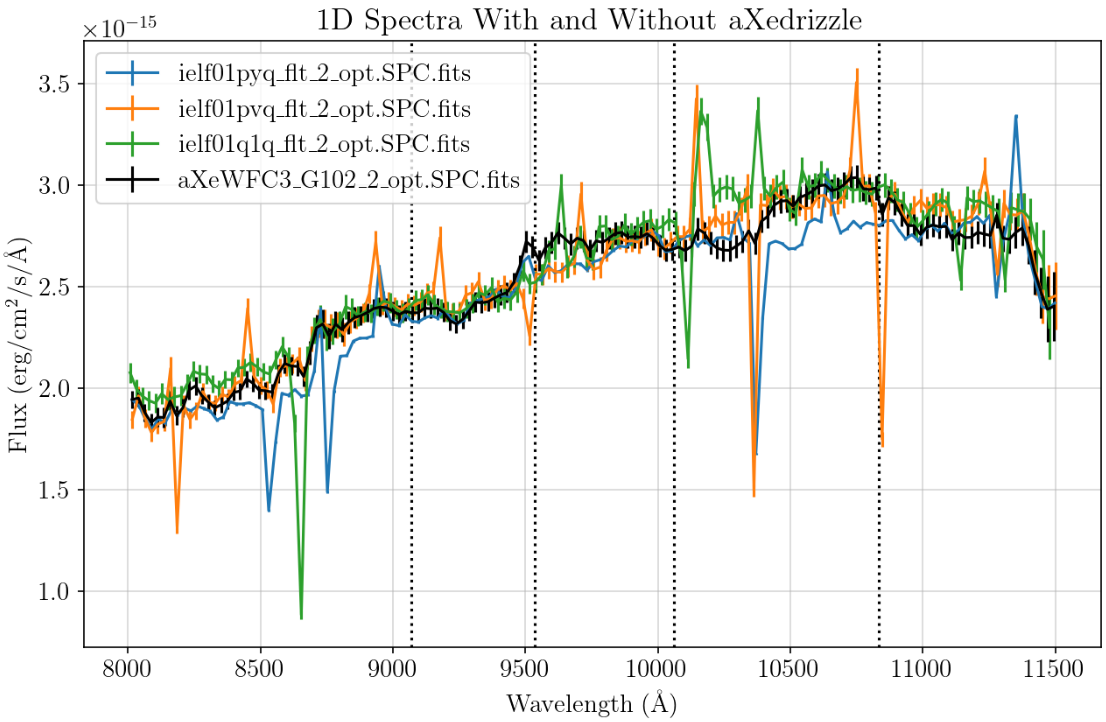

Figure 7: Spectra of a relatively faint source from the example data provided in the WFC3/IR cookbook. Spectra from three individual grism images are compared to a drizzled spectrum, in black. Note the absence of outliers (the sharp 'spike' features coming from bad pixels/cosmic rays) in the drizzled spectrum. The data in this figure come from WFC3 Calibration Program 16582; PI: Debopam Som

7 Conclusion

The `HSTaXe` cookbooks are a collection of six Jupyter Notebooks that explain and demonstrate how to preprocess ACS and WFC3 slitless spectroscopic data as well as extract 1D spectra with `HSTaXe`. In addition to describing the installation process and general workflow, we detail three major preprocessing steps that lead to optimally extracted spectra: a multi-component background subtraction for WFC3/IR grism data, an embedding procedure for all subarray data, and verifying that all exposures (dispersed and direct) have the same active WCS.

Like other tutorials and software demonstrations, these cookbooks are not meant to be a "one-size-fits-all" procedure. While there are common processing steps all HST grism users should follow when using `HSTaXe`, each cookbook is set up to use the provided example data. Users attempting to analyze their own data with these cookbooks should first become familiar with `HSTaXe` and its capabilities and limitations. We encourage users to read the Instrument Science Reports from the ACS[18] and WFC3[19] teams that explain the characteristics and

[18] ACS ISRs: https://www.stsci.edu/hst/instrumentation/acs/documentation/instrument-science-reports-isrs

[19] WFC3 ISRs: https://www.stsci.edu/hst/instrumentation/wfc3/documentation/instrument-science-reports-isrs

calibrations of the grism/prism observing modes. Additionally, the `HSTaXe` documentation page extensively covers installation, configuration files, file formats, and more. The `HSTaXe` software and cookbooks will continue to be maintained by STScI staff and any issues with the installation or running of the cookbooks should be reported to the HST General Help Desk. Finally, users seeking additional information and resources pertaining to slitless spectroscopic modes should visit the ACS prism/grism performance page and/or the WFC3 grism resources page.

Acknowledgments

The authors would like to thank Mariarosa Marinelli, Jenna Ryon, David Stark, and Joel Green for their thorough reviews and helpful feedback, which improved the quality of the report. We extend our gratitude to all those in the Data Management Division, especially Richard O'Steen and Duy Nguyen, who helped develop and troubleshoot the `HSTaXe` software. We are thankful to the HST Mission Office for their support of `HSTaXe` and for providing the necessary resources to complete this project. We would like to acknowledge the ACS and WFC3 Branch Managers Norman Grogin and Sylvia Baggett for their leadership and thoughtful guidance throughout this work.

References

Atek, H. et al. (Nov. 2010). "The WFC3 Infrared Spectroscopic Parallel (WISP) Survey". In: 723.1, pp. 104–115. DOI: 10.1088/0004-637X/723/1/104. arXiv: 1005.4068 [astro-ph.CO].

Bertin, E. and S. Arnouts (June 1996). "SExtractor: Software for source extraction." In: 117, pp. 393–404. DOI: 10.1051/aas:1996164.

Bohlin, R., S. E. Deustua, and N. Pirzkal (June 2015). *IR Grism Wavelength Solutions Using the Zero Order Image as the Reference Point*. Instrument Science Report WFC3 2015-10, 30 pages.

Bushouse, H. (Dec. 2008). *WFC3 IR Ground P-Flats*. WFC3 Instrument Science Report 2008-28, 14 pages.

Freudling, W. et al. (Nov. 2008). "The Hubble Legacy Archive NICMOS grism data". In: 490.3, pp. 1165–1179. DOI: 10.1051/0004-6361:200810282. arXiv: 0809.2465 [astro-ph].

Hack, Warren J. et al. (Aug. 2023). *spacetelescope/drizzlepac: Drizzlepac v3.6.2 release candidate 0*. Version 3.6.2rc0. DOI: 10.5281/zenodo.8264540. URL: https://doi.org/10.5281/zenodo.8264540.

Kluyver, Thomas et al. (2016). "Jupyter Notebooks - a publishing format for reproducible computational workflows". In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by Fernando Loizides and Birgit Scmidt. IOS Press, pp. 87–90. URL: https://eprints.soton.ac.uk/403913/.

Kümmel, M., H. Kuntschner, et al. (Sept. 2009). "The ST-ECF ACS Grism Hubble Legacy Archive Project". In: *Astronomical Data Analysis Software and Systems XVIII*. Ed. by D. A. Bohlender, D. Durand, and P. Dowler. Vol. 411. Astronomical Society of the Pacific Conference Series, p. 430.

Kümmel, M., J. R. Walsh, et al. (Jan. 2009). "The Slitless Spectroscopy Data Extraction Software aXe". In: 121.875, p. 59. DOI: 10.1086/596715. arXiv: 0812.1434 [astro-ph].

Mack, J. et al. (Aug. 2022). *Improved Absolute Astrometry for ACS and WFC3 Data Products*. Instrument Science Report ACS 2022-03, 37 pages.

Ogaz, S. and E. Tollerud (Nov. 2018). *Removing the Institute's Dependence on IRAF (You can do it too!)* STScI Newsletter Volume 35 Issue 03.

Pirzkal, N., A. J. Burgasser, et al. (Apr. 2009). "Spectrophotometrically Identified Stars in the PEARS-N and PEARS-S Fields". In: 695.2, pp. 1591–1603. DOI: 10.1088/0004-637X/695/2/1591. arXiv: 0901.3321 [astro-ph.GA].

Pirzkal, N., A. Pasquali, and M. Demleitner (July 2001). "Extracting ACS Slitless Spectra with aXe". In: *Space Telescope European Coordinating Facility Newsletter* 29, p. 5.

Pirzkal, N. and R. Ryan (Apr. 2020). *The Dispersed infrared background in WFC3 G102 and G141 observations*. Instrument Science Report WFC3 2020-4, 32 pages.

Pirzkal, N., C. Xu, et al. (Oct. 2004). "GRAPES, Grism Spectroscopy of the Hubble Ultra Deep Field: Description and Data Reduction". In: 154.2, pp. 501–508. DOI: 10.1086/422582. arXiv: astro-ph/0403458 [astro-ph].

Sosey, Megan et al. (July 2023). *spacetelescope/hstaxe: v1.0.5*. Version v1.0.5. DOI: 10.5281/zenodo.8136948. URL: https://doi.org/10.5281/zenodo.8136948.