PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Cure-WISE: HETDEX data reduction with Astro-WISE

Snigula, J., Cornell, M., Drory, N., Fabricius, Max., Landriau, M., et al.

J. M. Snigula, M. E. Cornell, N. Drory, Max. Fabricius, M. Landriau, G. J. Hill, K. Gebhardt, "Cure-WISE: HETDEX data reduction with Astro-WISE," Proc. SPIE 8451, Software and Cyberinfrastructure for Astronomy II, 845125 (24 September 2012); doi: 10.1117/12.925179

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2012, Amsterdam, Netherlands

Cure-WISE: HETDEX data reduction with Astro-WISE

J. M. Snigula[a,b], M. E. Cornell[c], N. Drory[d], M. Fabricius[a,b], M. Landriau[a,b], G. J. Hill[c], K. Gebhardt[c]

[a]Max Planck Institute for extraterrestrial Physics, PO Box 1312, Giessenbachstr., 85741 Garching, Germany;
[b]Unisternwarte München, Scheinerstr. 1, 81679, München, Germany;
[c]The University of Texas, 2515 Speedway, Stop C1402, Austin, Texas 78712-1206, USA;
[d]Instituto de Astronomía, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Ciudad Universitaria, C.P. 04510 México D.F., México

ABSTRACT

The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) is a blind spectroscopic survey to map the evolution of dark energy using Lyman-alpha emitting galaxies at redshifts $1.9 < z < 3.5$ as tracers. The survey instrument, VIRUS, consists of 75 IFUs distributed across the 22-arcmin field of the upgraded 9.2-m HET. Each exposure gathers 33,600 spectra. Over the projected five year run of the survey we expect about 170 GB of data per night. For the data reduction we developed the Cure pipeline. Cure is designed to automatically find and calibrate the observed spectra, subtract the sky background, and detect and classify different types of sources. Cure employs rigorous statistical methods and complete pixel-level error propagation throughout the reduction process to ensure Poisson-limited performance and meaningful significance values. To automate the reduction of the whole dataset we implemented the Cure pipeline in the Astro-WISE framework. This integration provides for HETDEX a database backend with complete dependency tracking of the various reduction steps, automated checks, and a searchable interface to the detected sources and user management. It can be used to create various web interfaces for data access and quality control. Astro-WISE allows us to reduce the data from all the IFUs in parallel on a compute cluster. This cluster allows us to reduce the observed data in quasi real time and still have excess capacity for rerunning parts of the reduction. Finally, the Astro-WISE interface will be used to provide access to reduced data products to the general community.

Keywords: Astro-WISE, Integral field spectrograph, VIRUS, HETDEX, data reduction, dark energy, survey

1. HETDEX

The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) is a blind spectroscopic survey to map the evolution of dark energy using Lyman-alpha emitting galaxies as tracers. The survey will use an array of 75 integral field spectrographs (IFSs) on the upgraded 9.2 m HET[1,2] called the Visible Integral field Replicable Unit (IFU) Spectrograph (VIRUS[3]). Each of these instruments consists of an IFU with 448 1.5 arcsec fibers that feed a pair of spectrographs with a fixed bandpass of $360 - 540$ nm corresponding to a redshift range of $1.9 < z < 3.5$ and a resolving power $R \sim 700$. Each exposure gathers 33,600 spectra. The IFUs are designed such that three dithers fill the complete area of an IFU. With an observing time of 20 minutes per field we expect to reach a line flux of 3.5×10^{-17}ergs s^{-1}cm^{-2} and a limiting magnitude of $m_{AB} \sim 22$.

The survey will consist of a main field covering $300\,\mathrm{deg}^2$ centered at 13^h, $+53°$ in a $42 \times 7\,\mathrm{deg}^2$ layout. The area will be observed with a 1/4.5 fill factor, resulting in an area of 60 deg^2 covered by spectra. This field will be complemented by an equatorial field centered at 1.5^h, 0° covering an additional $28 \times 5\,\mathrm{deg}^2$ overlapping with many other surveys (e.g. DES, Spitzer, Herschel, HSC, SDSS). Observations of these will be finished within three years. After the completion of the main survey, HETDEX plans to extend the areas covered to $42 \times 10\,\mathrm{deg}^2$ and $28 \times 8\,\mathrm{deg}^2$ respectively. The extensions should be observed within another two years.

The baseline survey will deliver spectra of 800,000 LAEs in a 6 Gpc^3 volume with $1.9 < z < 3.5$ and one million [OII] emitters with redshifts up z = 0.48. The survey will in addition cover 400,000 other galaxies, 250,000 stars, 2000 galaxy clusters, 7000 QSOs with $z < 3.5$, 5000 local galaxies with resolved kinematics data and 20,000 NVSS radio sources.

Software and Cyberinfrastructure for Astronomy II, edited by Nicole M. Radziwill, Gianluca Chiozzi,
Proc. of SPIE Vol. 8451, 845125 · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.925179

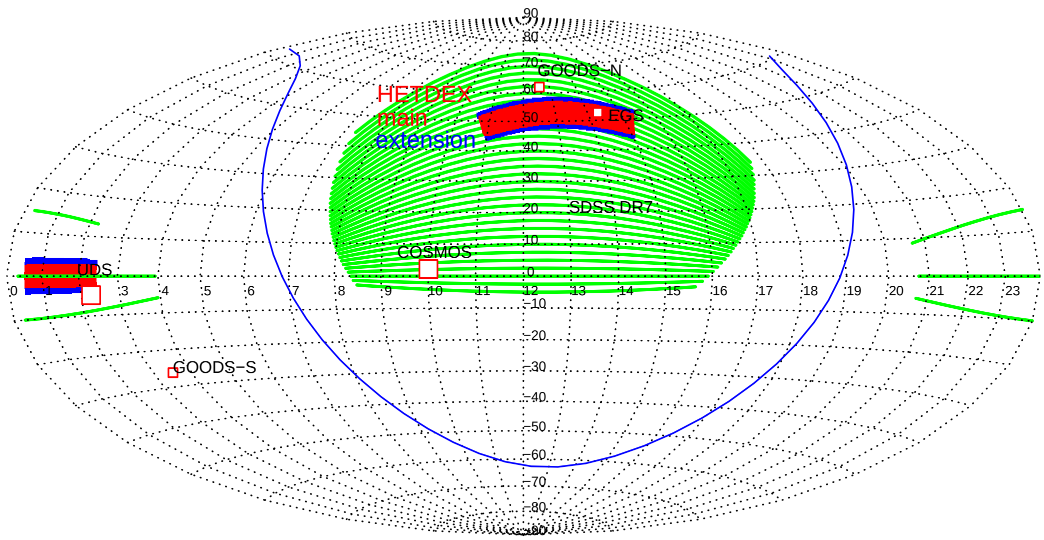

Figure 1. Main HETDEX three year survey field, equatorial field, and the proposed extensions.

2. CURE - THE HETDEX REDUCTION PIPELINE

Given the size of the HETDEX survey we need a specialized reduction pipeline capable of taking care of the expected data volume within a real-time timeframe, i.e. reducing the data faster than new data arrives. The whole process has to be sufficiently robust, since the number of observed frames renders the manual inspection of all images impossible. One key feature of the pipeline is the detailed error propagation for each observed pixel. For the basic reduction steps we use the fitstools package developed by Gössl and Riffeser.[4]

A typical HETDEX dataset consists of a set of bias frames; a set of flat frames, i.e. observations of the twilight sky, needed for the automated detection of the individual spectra on the chip and the calibration of the throughput of the individual fibers; a set of observations of calibration emission line spectra for the wavelength calibration; and the actual science frames. All images are bias and overscan corrected and an initial error frame using the photon noise as an initial value is created. The twilight flat frames are averaged to a maximum signal to noise normalized image, the trace frame. From the calibration lamp spectra the kappa-sigma clipped median is calculated, providing an arc frame. After this basic image reduction, in the next step we measure in the arc and trace frame, the distortion of the spectra. The wavelength calibration and the measurements of the positions of the individual fiber positions across the detector is described by the *distortion model*. Further models describe the fiber profile in crossdispersion direction (the *fiber model*) and the shape of the instrumental Point Spread Function (PSF) as function of the position on the detector.

The background sky is subtracted from each science frames after the basic reduction steps. The proper execution of this step is especially important for a blind survey like HETDEX looking for faint emission line galaxies. We first reject fibers with continuum sources based on medium filtering technique. The sky spectrum is then fitted using a spline within a window in crossdispersion direction. The use of a window increases the signal to noise, while still avoiding the mixing of regions with significantly different spectral resolution. This noise free model of the sky background is then subtracted from the science image using the known distortion and fiber models. The final step of the pipeline is the automated detection of the sources in the science frames. In the case of HETDEX it is more difficult to detect sources than it would be for an imaging survey, since one point-like source could appear in different positions in the image, due to the fiber based IFUs, and on different positions for each image of a target, due to the dithering between observations. We follow a quasi-Baysian approach and place imaginary point sources on the sky. We use the previously determined distortion-, fiber and PSF models to compute how such a source would appear on the CCDs and then compare this to the actually observed counts at the corresponding detector positions. By stepping the positions of the imaginary source both in RA and DEC and also in wavelength, we obtain a three dimensional map of the significance of a detection. In the next step the significance map is segmented, i.e. searched for consecutive areas exceeding a certain threshold.

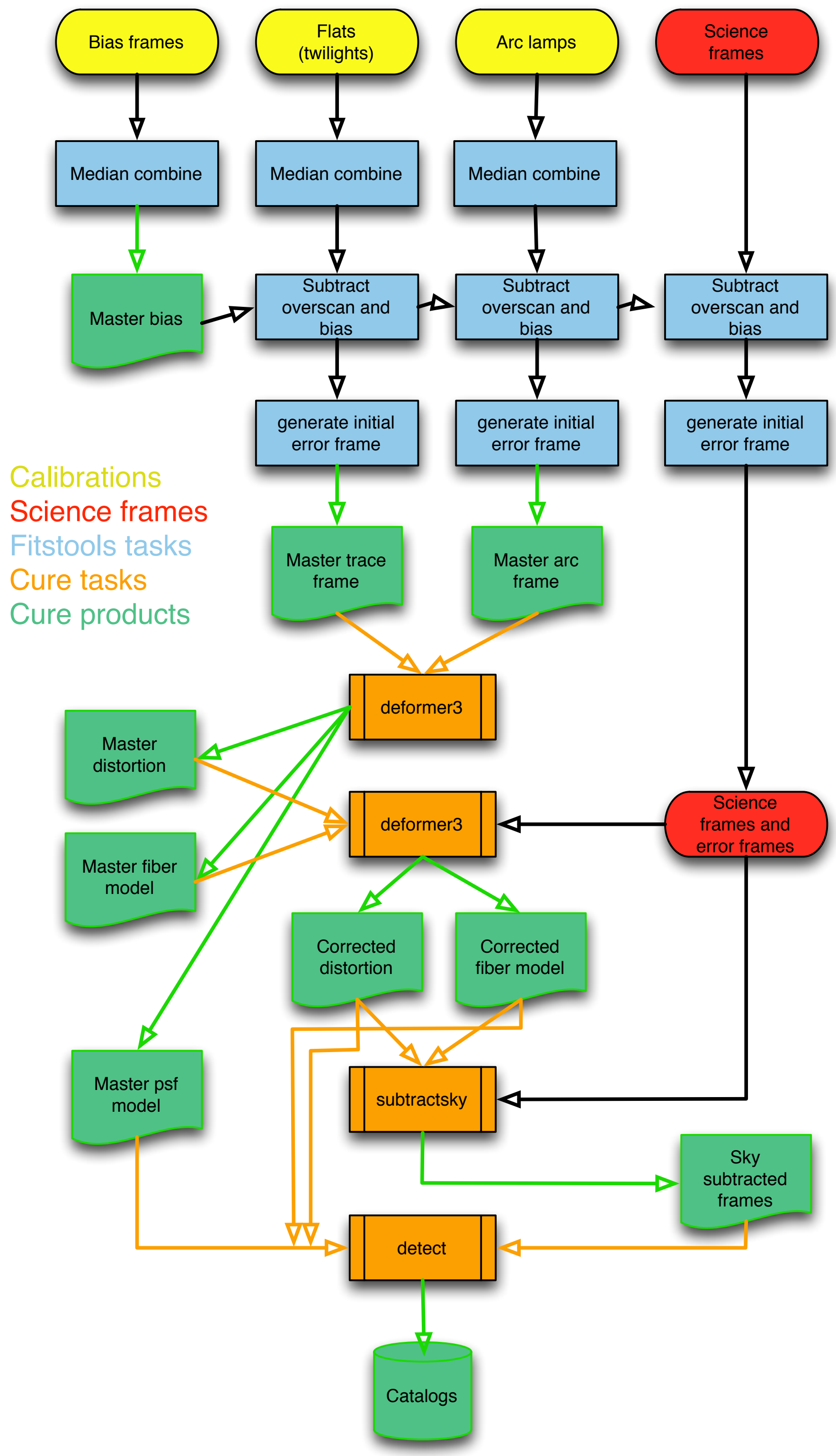

Figure 2. Flowchart of Cure data reduction pipeline. Shown are the input images at top, the basic reduction pipeline with the fitstools package, the Cure specific tasks and the various output products.

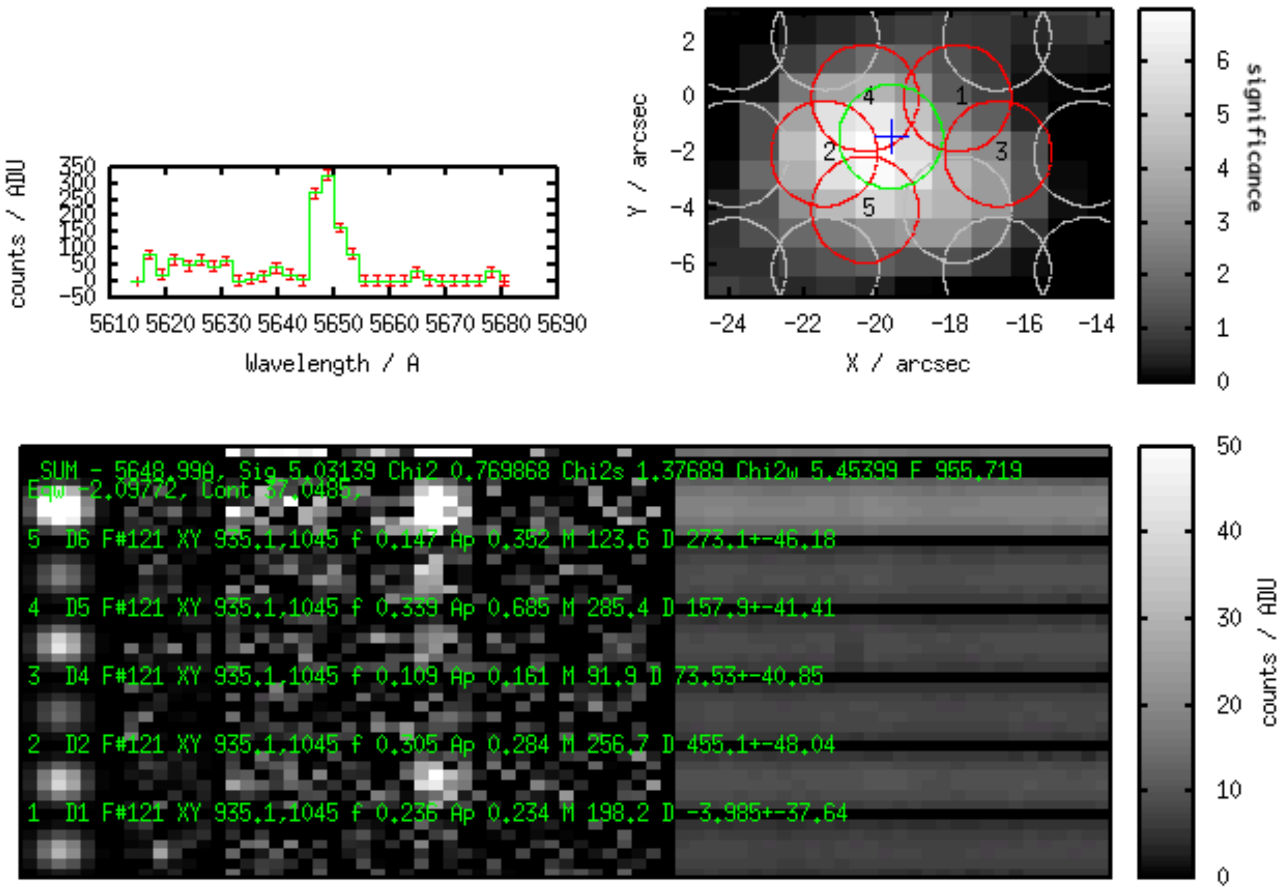

Figure 3. Emission line detection diagnostic plot. The plot shows a 1-D extraction of the line (top-left), a 2-D image of the segment with fibers overlayed (top right). All fiber positions surrounding the detected line are shown, the contributing fibers, and the detection aperture. Note that this is NOT an on-sky reconstructed flux image. This is the significance map - the Bayesian significance of a detection assuming a source at a given position, hence the much smaller than the fiber sampling. You can think of this as a likelihood map. Finally, the bottom plot shows the error images (right), the science images (middle), the chi2-map, and the model PSF (left) for each contributing fiber (bottom rows) and the sum all these (top row).

Based on the morphology of the found segments we can differentiate between various types of objects, e.g. nearby galaxies would be extended in all three dimensions, a star would be point like in RA and DEC, but elongated in wavelength and an Lyman-alpha emitter would be point-like in all three dimensions. To erase possible false detections (like cosmic rays), for all candidates we compare the pixel-level flux distribution on the CCD with the expectation from the detailed modeling of the light propagation through the spectrograph including the instrumental point spread function.

We created various tools for the visualization of the detected sources (see Fig. 3 for an example), and for the inspection and quality control of the pipeline products, e.g. to check the form of the reconstructed fiber model.

3. CURE-WISE

We expect on a typical night to observe about 24 fields with the 75 IFUs resulting in an average nightly data volume (including calibration data) of approximately 170 GB. This data rate calls for a highly automated and, most importantly, parallelized reduction approach.

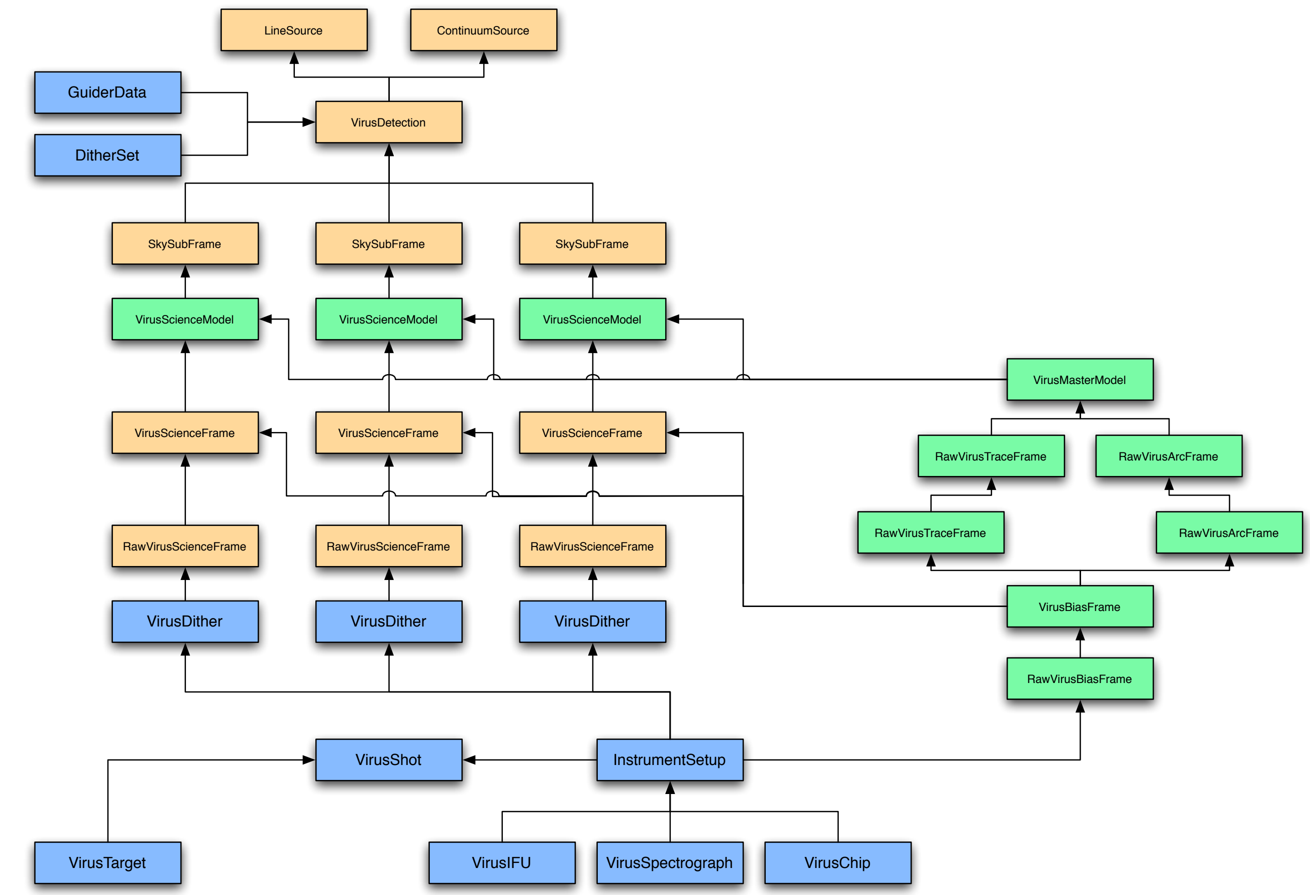

Figure 4. Flowchart of Cure-WISE pipeline. Shown are the different Astro-WISE python objects encapsulating the Cure pipeline shown in Fig. 2, and their dependency tree.

We decided to use Astro-WISE[5] (http://www.astro-wise.org), for the parallel reduction, the management and bookkeeping of the observed data (we expect more than 35000 individual FITS images on a typical night) and the organization and publication of the detected sources. Another benefit of using Astro-WISE is the dependency tracking it provides for the whole reduction process. This makes it possible to retrace the reduction down to the individual frames for each detected source. Astro-WISE stands for Astronomical Wide-field Imaging System for Europe. It is an environment consisting of hardware and software which can be federated between different institutes. In our case the federated approach gives all users of the HETDEX consortium equal access to the database and the data reduction facilities. A typical Astro-WISE node consists of one or more ORACLE database servers responsible for the metadata about the images, data servers storing the actual images and a compute cluster for the parallel reduction. A user can of course also use an Astro-WISE node remotely, without the need for a local database or data server. Astro-WISE by default supports various imagers and can be adapted to new instruments as well.

For HETDEX we created python wrappers for all the reduction steps in the Cure pipeline (Cure-WISE), encapsulating the it in Astro-WISE. With this approach we gain complete management of all the observed data. The wrappers give us access to the Astro-WISE batch systems for parallel data reduction of all IFUs at once. The system is designed for automated quality control checks at various steps in the pipeline, providing us with means to reduce the HETDEX data unsupervised. According to our first tests with mock HETDEX data, we estimate that we will be able to reduce the observations not only in real time, but also retain additional processing power for re-reduction of the data with improved code and for analysis by the scientists. The database

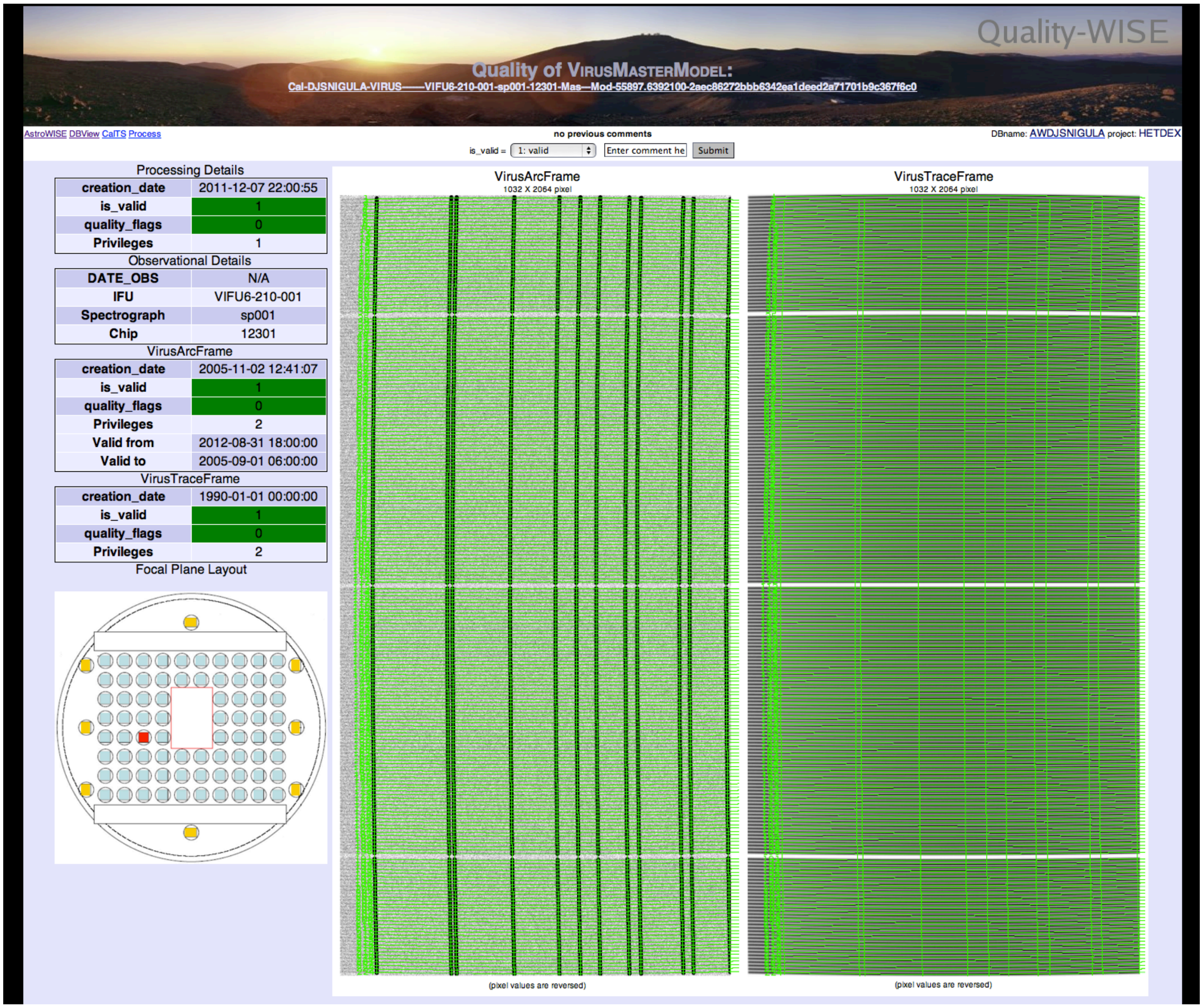

Figure 5. An example quality control output of Cure-WISE, showing basic processing data for a distortion and fiber model, as well as inspection plots for the tracing of the fibers on the chip (left) and the wavelength solution (right).

backend of Astro-WISE allows us to quickly and efficiently search and create subsets of the detected LAEs or combine different detections to more complex objects, e.g. a galaxy with line emission in the spiral arms. Since Astro-WISE is designed to be used with many instruments, we can use it to reduce the accompanying imaging survey and match the resulting catalog against our detections.

We adopted the web services provided by Astro-WISE for quality control and access to the datasets. We plan public data releases to the community as a Year 1 and a Year 1-3 data release. Each data release will consist of all flux-calibrated spectra, their absolute positions on the sky to within one arcsec, and a timestamp. For each spectrum, we will provide an estimate of the spatial and spectral point-spread function, and the flux limit including transparency (to 5% accuracy). The flux calibration will be accurate to at least 10% at each wavelength, with a goal of 5% accuracy. Data access will be provided via a web interface in the Astro-Wise system.

ACKNOWLEDGMENTS

HETDEX is run by the University of Texas at Austin McDonald Observatory and Department of Astronomy with participation from the Ludwig-Maximilians-Universität München, Max-Planck-Institut für Extraterrestrische-Physik (MPE), Leibniz-Institut fuer Astrophysik Potsdam (AIP), Texas A&M University, Pennsylvania State University, Institut für Astrophysik Göttingen, and University of Oxford. In addition to Institutional support, HETDEX is funded by the National Science Foundation (grant AST-0926815), the State of Texas, the US Air Force (AFRL FA9451-04-2-0355), and generous support from private individuals and foundations.

REFERENCES

[1] Ramsey, L. W., Adams, M. T., Barnes, T. G., Booth, J. A., Cornell, M. E., Fowler, J. R., Gaffney, N. I., Glaspey, J. W., Good, J. M., Hill, G. J., Kelton, P. W., Krabbendam, V. L., Long, L., MacQueen, P. J., Ray, F. B., Ricklefs, R. L., Sage, J., Sebring, T. A., Spiesman, W. J., and Steiner, M., "Early performance and present status of the Hobby-Eberly Telescope," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], Stepp, L. M., ed., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **3352**, 34–42 (Aug. 1998).

[2] Savage, R., Booth, J., Cornell, M., Good, J., Hill, G. J., Lee, H., MacQueen, P., Rafal, M., Vattiat, B., Gebhardt, K., Beno, J., Zierer, J., Perry, D., Rafferty, T., Ramiller, C., Taylor, III, C., Beets, T., Hayes, R., Heisler, J., Hinze, S., Soukup, I., Jackson, J., Mock, J., Worthington, M., Mollison, N., Molina, O., South, B., Wardell, D., and Wedeking, G., "Current status of the Hobby-Eberly Telescope wide-field upgrade," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **7733** (July 2010).

[3] Hill, G. J., Lee, H., Vattiat, B. L., Adams, J. J., Marshall, J. L., Drory, N., Depoy, D. L., Blanc, G., Bender, R., Booth, J. A., Chonis, T., Cornell, M. E., Gebhardt, K., Good, J., Grupp, F., Haynes, R., Kelz, A., MacQueen, P. J., Mollison, N., Murphy, J. D., Rafal, M. D., Rambold, W. N., Roth, M. M., Savage, R., and Smith, M. P., "VIRUS: a massively replicated 33k fiber integral field spectrograph for the upgraded Hobby-Eberly Telescope," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **7735** (July 2010).

[4] Gössl, C. A. and Riffeser, A., "Image reduction pipeline for the detection of variable sources in highly crowded fields," *A&A* **381**, 1095–1109 (Jan. 2002).

[5] Valentijn, E. A., McFarland, J. P., Snigula, J., Begeman, K. G., Boxhoorn, D. R., Rengelink, R., Helmich, E., Heraudeau, P., Verdoes Kleijn, G., Vermeij, R., Vriend, W.-J., Tempelaar, M. J., Deul, E., Kuijken, K., Capaccioli, M., Silvotti, R., Bender, R., Neeser, M., Saglia, R., Bertin, E., and Mellier, Y., "Astro-WISE: Chaining to the Universe," in [*Astronomical Data Analysis Software and Systems XVI*], Shaw, R. A., Hill, F., and Bell, D. J., eds., *Astronomical Society of the Pacific Conference Series* **376**, 491 (Oct. 2007).