Modeling and removing contamination in WISP grism data

Avery Garon
University of Minnesota – Twin Cities

Introduction

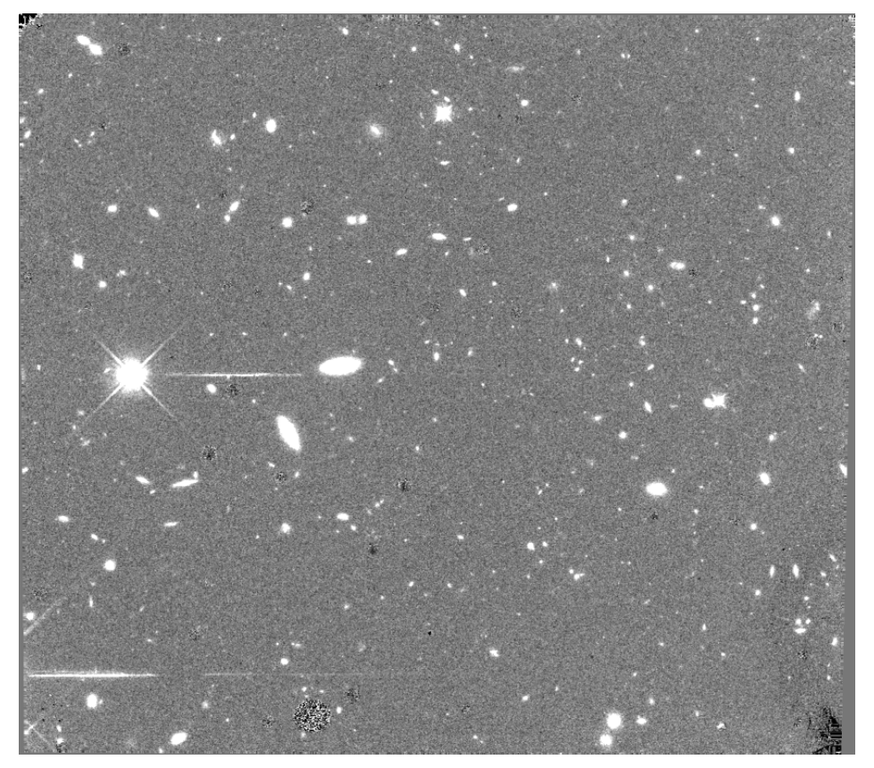
The study of galaxies in the early universe is necessary for understanding how modern galaxies evolved. The bulk of star formation occurred during $0.5<z<2.5$ (e.g. Daddi et al. 2007), or roughly 4.5 to 9.2 billion years ago. Probing this epoch and earlier requires infrared spectroscopy, due to the redshifting of important spectral features at those ages.

The WFC3 Infrared Spectroscopic Parallel (WISP) Survey (Atek et al. 2010) uses the WFC3 grated prisms ("grisms") on the Hubble Space Telescope to measure the 2D spectra of fields of galaxies. As a space-based survey, WISP can avoid contamination from the atmosphere, letting it resolve faint galaxies at a wide range of wavelengths. On the other hand, the slitless nature of the survey does not prevent the spectra of different objects from overlapping.

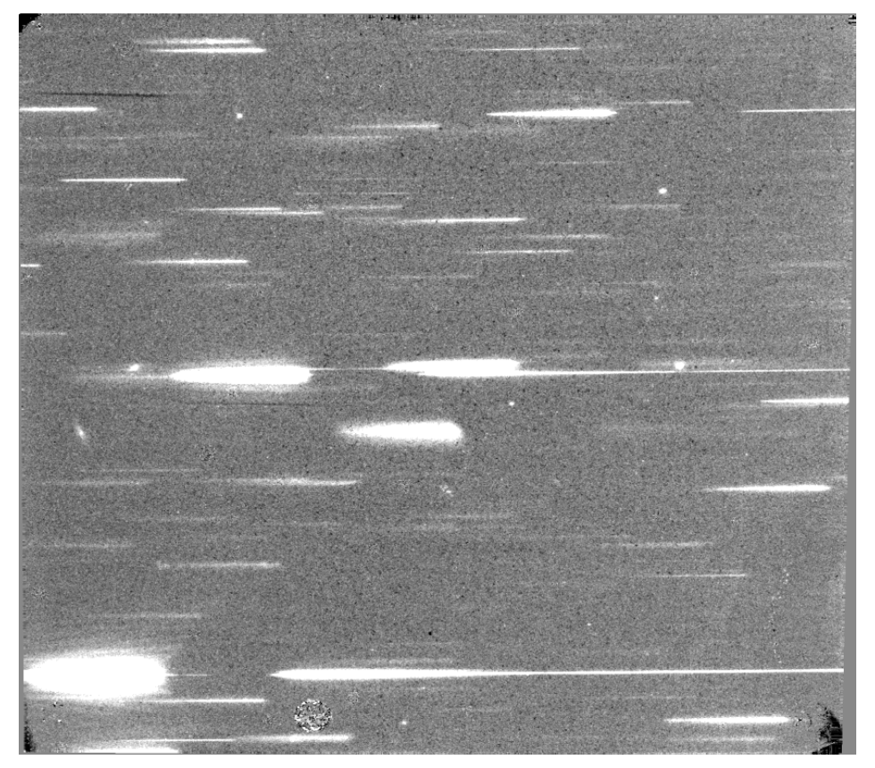
The purpose of this project is to create an algorithm that can model and remove the overlapping spectra. This will allow for better analysis of redshifted continuum-emission features, like the 4000 Å and Balmer breaks. This is especially important for studying quiescent galaxies at high redshifts, which do not have emission lines (Bedregal et al. 2013).

Data products

This algorithm uses three types of data products from WISP: direct images, dispersed images, and object catalogs. The direct images are taken with the F110W and F140W filters of WFC3, which cover wavelength ranges of 900 to 1400 nm and 1200 to 1600 nm. The images are dispersed using the G102 and G141 grisms, which correspond to the same wavelength ranges (Dressel 2016). The object catalogs contain information that includes position, morphology, and magnitude for all objects in the images.



Direct F110W image

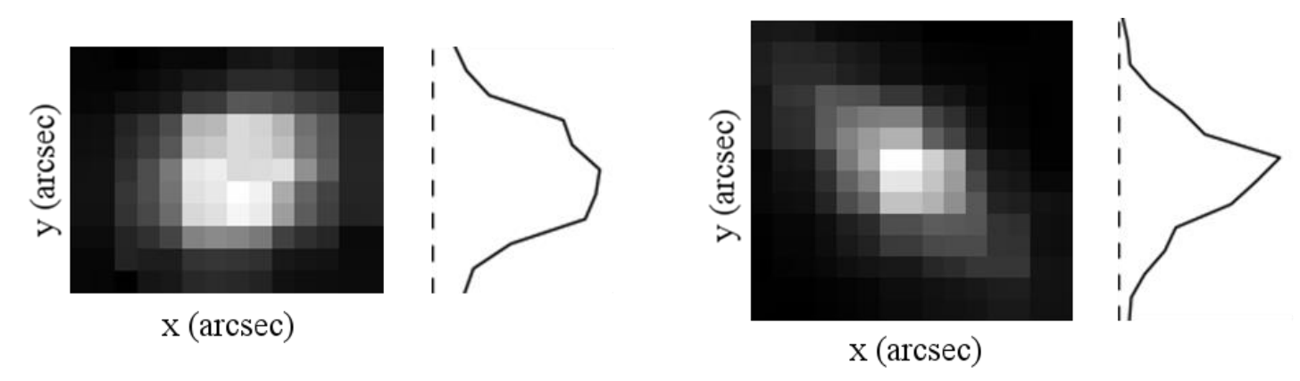


Dispersed G102 image

Process

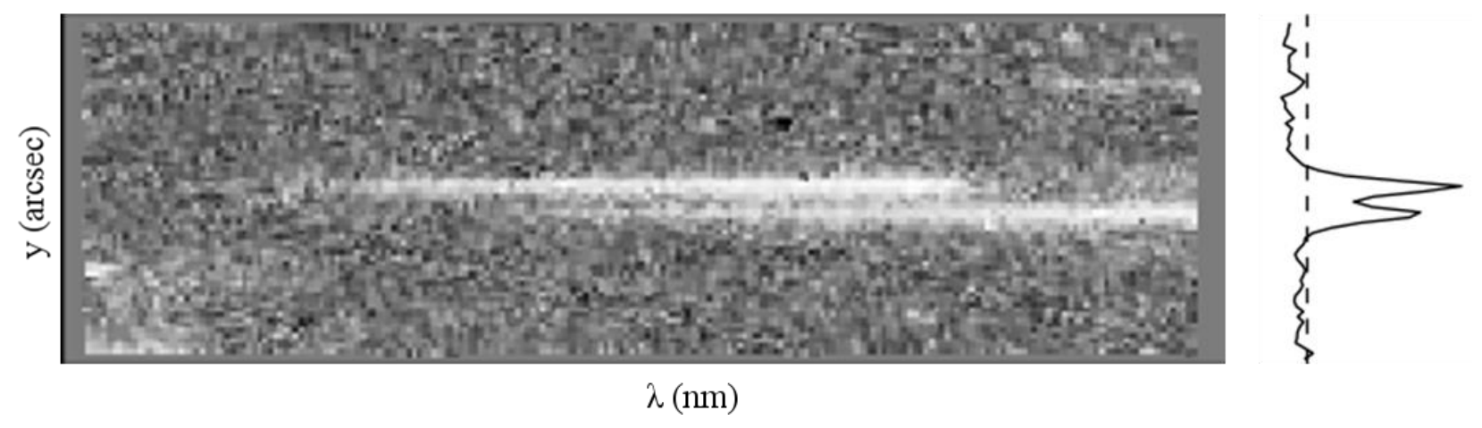
1. Isolate individual objects

   Each galaxy must be isolated from the direct and dispersed images. The object catalog contains information describing the projected shape of the direct image, so stamps the size of the ellipse are cut from the direct image. Similar cutouts of the 2D spectra are provided by WISP.



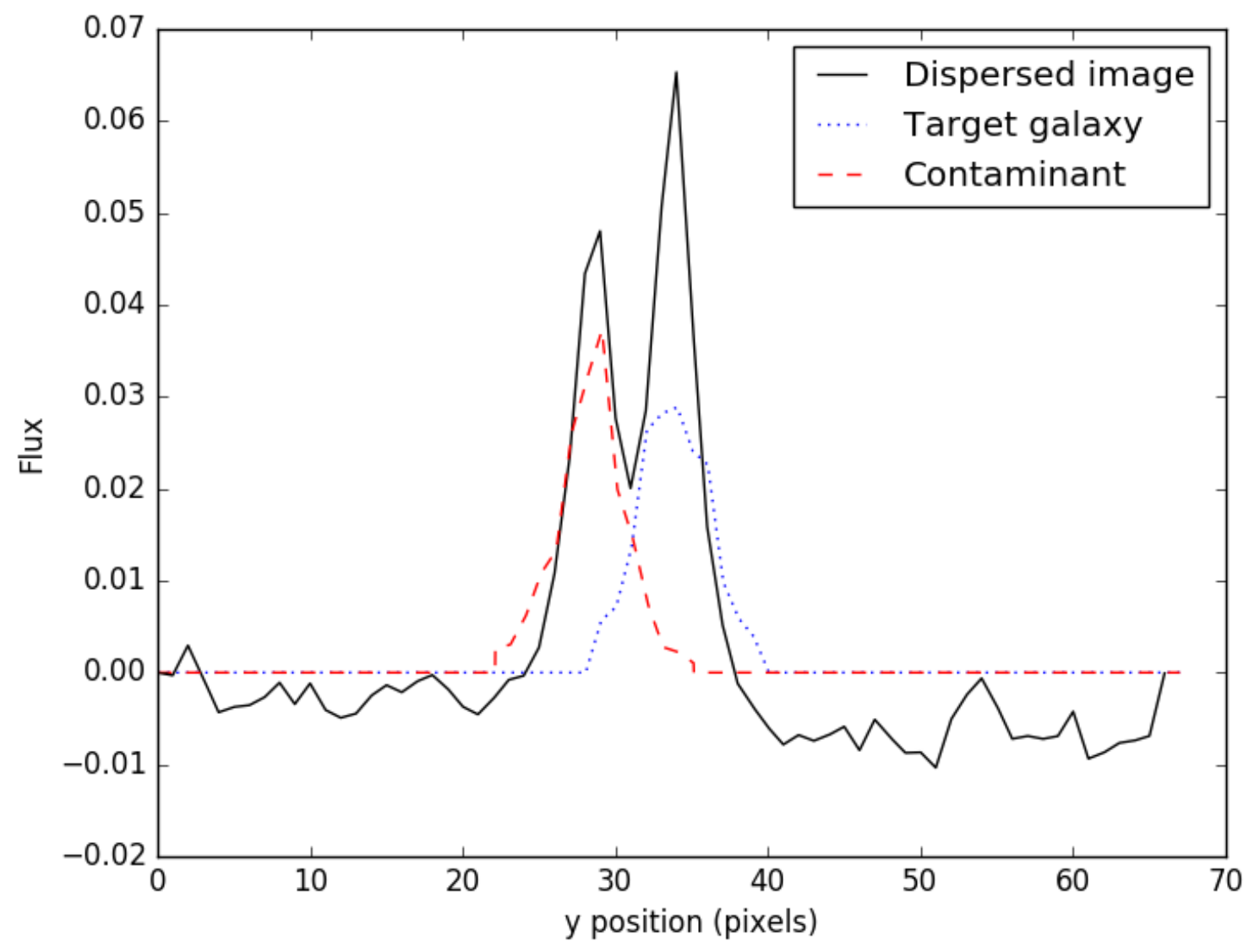
2. Determine sources of contamination

   I estimate the width of each galaxy's 2D spectrum to be equal to the wavelength range of the grism plus the extent of the galaxy itself in the direct image. Because the contaminants are offset in the *x*-direction in space, when the grism disperses the image the contaminants become offset in the λ-direction. For example, a contaminant physically located to the right of the target galaxy in the field cannot be contaminating the leftmost part of the grism data. For an object to contaminate the target spectrum, its width must overlap the width of the target spectrum. In the *y*-direction, the image of the galaxy is not dispersed, so the possible contaminant must physically overlap the range of *y*-coordinates spanned by the galaxy.



3. Create spatial profiles

   The intensity profiles of each galaxy are derived from the stamps by taking the middle third of the galaxy's stamp in the *x*-direction and collapsing it in the *y*-direction. Because the grism does not disperse the image in the *y*-direction, this profile has the same shape as the spatial profile of the 2D spectrum. When the individual spectra are extracted from the larger image, they are centered on the object of interest, which may not lie exactly on a pixel. Because of this, there may be an offset (of less than one pixel) between the "true" center of the galaxy and the center of the cropped grism image. To counteract this, the profiles are interpolated using `interp1d`, the one-dimensional linear interpolation method provided by SciPy, and allowed to vary in position slightly.



4. Fit the model globally to find offsets

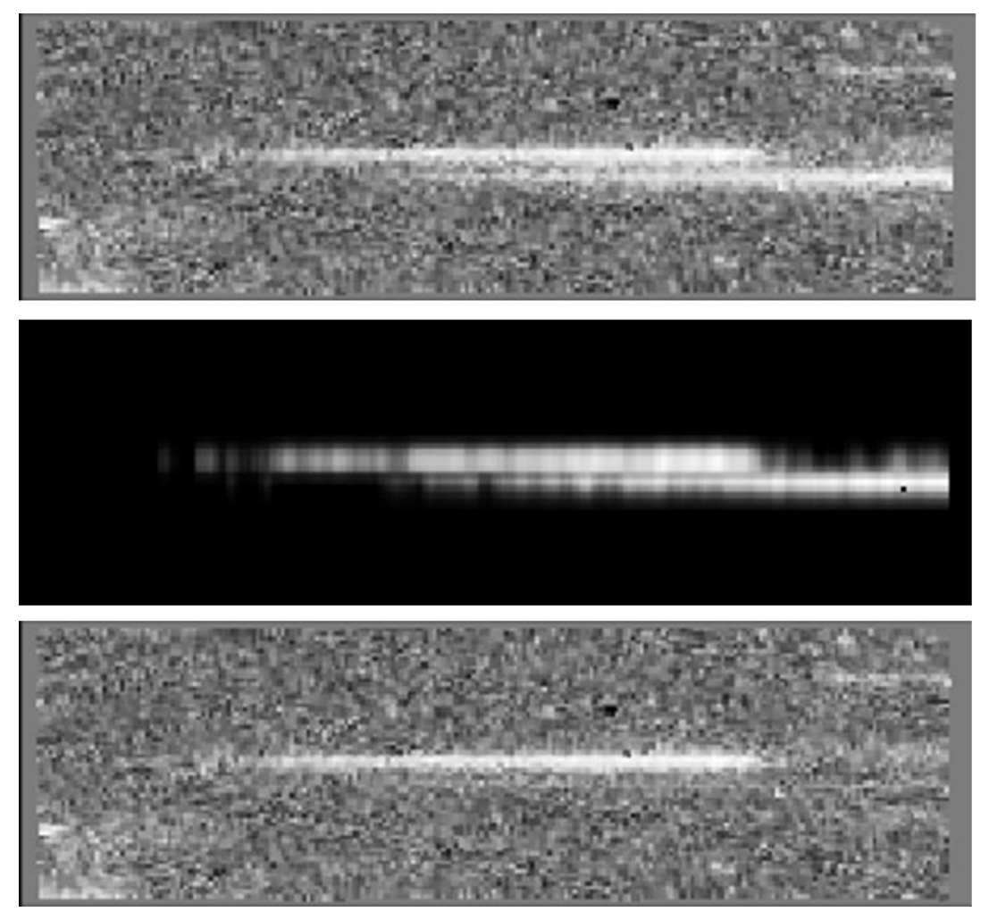
   The spatial intensity profile of the dispersed image is a linear combination of the profiles of each object supplying flux to it. There are two parameters I vary for each profile when trying to fit the model: the amplitude and the subpixel offset. I first fit the profiles to the collapsed grism data. I find the optimum fit via χ^2-minimization, varying those parameters for each profile. The interpolated profiles are evaluated at a 0.01 pixel scale when comparing the model to the grism data, which sets the precision for the subpixel offsets.

5. Fit the model locally to find amplitudes

   At every wavelength in the grism data, the three pixel column centered at that wavelength (determined by the grism's resolution) is averaged to obtain the observed profile at that wavelength. The amplitude of the main galaxy's profile, as well as any contaminant that is contributing flux at that wavelength, is calculated by fitting the profiles to the averaged data column using χ^2-minimization, except this time the offsets are held constant and equal to the overall offset.

6. Subtract the contamination

   Once the amplitudes at each wavelength are known, I generate a two-dimensional model of the contamination. This model is generated by taking the individual profiles at each wavelength, weighting their amplitude as found by the fitting at each wavelength, offsetting them by the amount found by the fitting to the collapsed dispersed image, and summing them all together. This gives the expected profile at each wavelength. To clean the dispersed image, the modeled spectra of the contaminating galaxies are subtracted from the original image.

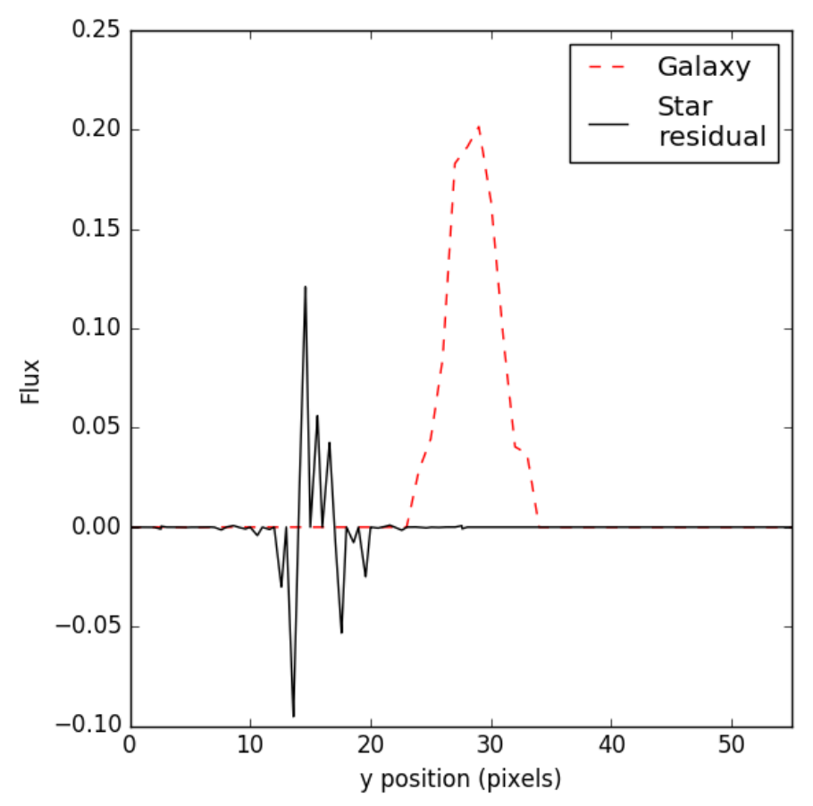


Original image

Contamination model

Decontaminated image

Discussion



This process can be applied to any WISP field and does not depend on wavelength. At this stage, only contamination from galaxies can be removed. Stars are very bright and peak very sharply, so the inaccuracy from the interpolation is similar to the total magnitude of the galaxy.

Another point for future improvement is incorporating the response curve of the grism when determining the wavelength extent of the contaminating spectra. The current implementation assumes a constant sensitivity over all wavelengths. An improved method will apply the response curve as a prior to improve fitting at the extremes of the wavelength range.

References

Atek, H., Malkan, M., McCarthy, P., et al. 2010, ApJ, 723, 104

Bedregal, A. G., Scarlata, C., Henry, A. L., et al. 2013, ApJ, 778, 126

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Dressel, L. 2016, "Wide Field Camera 3 Instrument Handbook, Version 8.0" (Baltimore: STScI)

Acknowledgments

I would like to thank Vihang Mehta for his assistance in using the WISP data and refining my code. I would also like to thank Claudia Scarlata, my advisor on this project, for guiding my progress towards a fruitful result.

Further information

This algorithm is publically available and will eventually be incorporated into the WISP pipeline. The source code is available on GitHub at `dx.doi.org/10.5281/zenodo.49092`.