

# HSRED v1.2: An Updated IDL Pipeline for Hectospec Data Reduction

Sean Moran, SAO Telescope Data Center

May 23, 2014

## Abstract

HSRED is an IDL-based data reduction pipeline for Hectospec, originally developed by Richard Cool (MMTO), and based upon code originally developed for SDSS. This report details the changes that have been made at the SAO Telescope Data Center (TDC) to produce version 1.2 of HSRED, and presents a number of key comparisons between the performance of the new pipeline and that of both the old version of HSRED, and the SPECROAD IRAF-based pipeline that is being superseded. Compared to SPECROAD, the new HSRED results in improvements in sky subtraction, adds an optional red fringing correction, implements correction for telluric A- and B-band absorption, models and corrects for the red leak, and improves the output data products. Compared to the old version of HSRED, we have improved accuracy and reliability of wavelength calibration (including support for 600-line data), ported superior cosmic-ray and fiber throughput-correction algorithms from SPECROAD, and implemented a simplified algorithm for co-adding exposures, in addition to the improved sky subtraction, red leak, and telluric correction mentioned above.

In each of the sections below, we summarize code changes and present relevant comparisons for each of these parts of the code. At the end, we also briefly discuss our model for integrating the code into our operations at TDC, and for distributing the code publicly.

## 1 Wavelength Calibration

Thanks to the long-term efforts of Susan Tokarz, wavelength solutions for both 270-line and most configurations of the 600-line grating were well-calibrated under the SPECROAD pipeline. In contrast, HSRED was only set up to deal with 270-line data, and its initial performance was mixed: due largely to a few coding bugs, small clusters of fibers with ‘missed’ wavelength solutions, especially at the red end, were common. On the other hand, the HSRED algorithm has greater flexibility to begin with a generic initial guess and then hone-in on the wavelength solution, compared to IRAF (“reidentify”), which required a new template wavelength solution be generated from time to time.

Attempting to merge the strengths of both methods, for HSRED 1.2 we adopt Susan’s refined HeNeAr line list for the IDL code, and use a similar 5th-order Legendre fit. Combining this with bug-fixes, we now recover nominal residuals for 270-line reductions similar to what the IRAF pipeline reports, about 0.08Å rms, with fits based on  $\sim 60$  identified lines. As with the IRAF pipeline, residuals may be somewhat higher at the blue end; Figure 1 shows the fit residuals for all matched lines in a single, typical fiber.

HSRED further tweaks wavelength solutions to match the measured positions of sky lines in every science frame, and so the ultimate accuracy at the red end is quite high (Typical shifts are  $< 0.1\text{\AA}$ ). Given the lack of sky lines in the blue, final accuracy at the blue end is the biggest question. To investigate, I compared old vs new pipeline for a number of science spectra with many blue lines. In Figure 2, top, we show a spectrum of standard star HD-192281, zoomed in to the blue end. Balmer-series absorption lines are prominent in this star, and the rest wavelengths of several are marked with vertical lines. By eye, no serious deviations between the expected and observed positions of the Balmer lines are apparent. In the lower panel, we plot the difference between measured and expected wavelength for each Balmer line. Black points are from the data reduced with HSRED 1.2, and red points are from data reduced with SPECROAD. Error bars indicate the uncertainty in the measured centroid of each line. Centroids were measured with

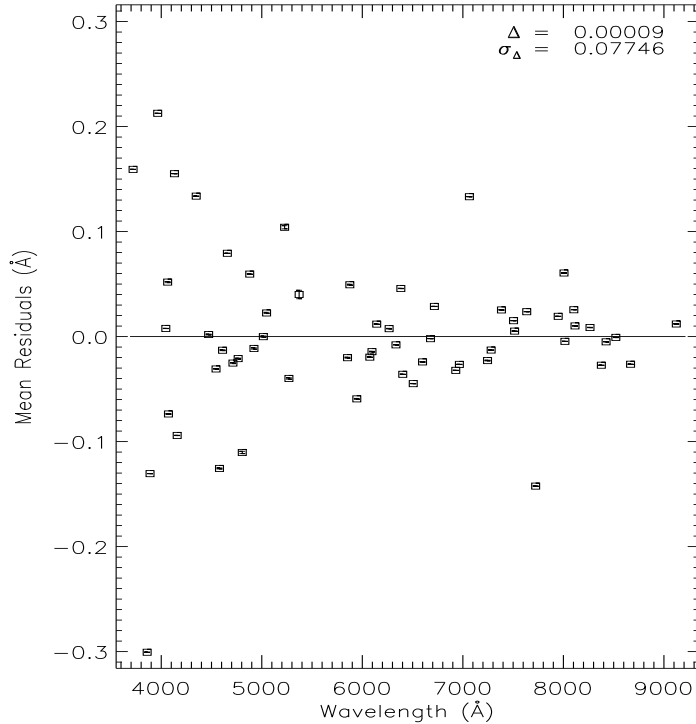


Figure 1: Wavelength fit residuals for a single, typical fiber observed with the 270-line grating. Boxes indicate the Fit–line center for each of the matched HeNeAr lines.

a Gaussian fit plus a low-order removal of the continuum shape. The green point at H-alpha is from the same HSRED spectrum as the black point, but with a slightly different choice of wavelength window for the Gaussian fit: clearly some of the scatter in this plot is due to the crudeness of these fits, and the large deviation at Halpha is likely not real. While in general the two pipelines produce indistinguishable results, it is also clear that the very blue end is still a challenge to calibrate accurately.

As mentioned above, reduction of 600-line data, for all typical central wavelengths, has been enabled in the IDL pipeline. Initial wavelength solutions and line lists generated by Susan (both HeNeAr and PenRay, where appropriate) were used as the baseline for the ‘initial guesses’ hard-coded into the IDL. Each configuration was then tuned and tested by hand until residuals of  $\sim 0.04\text{\AA}$  rms were achieved. The algorithm should be flexible enough to lock onto the right solution without the need for periodic updates of the initial guess, but this remains to be seen. Most solutions are solved for using 50-60 identified lines, though a few uncommon configurations (with less attention paid to refining the line list) are based on closer to  $\sim 30$  lines.

## 2 Flat Fielding: Fiber Throughput and Fringing

As a precursor to modeling and removing the red leak (and in order to improve sky subtraction), Igor Chilingarian suggested that it would be important to refine our flat-fielding performance. In particular, proper correction of red fringing would help with the red leak subtraction. In both SPECROAD and the original HSRED, flat fielding was handled in a similar fashion, with both dome flats and sky flats used to correct for pixel response and fiber throughput, respectively. Red fringing was divided out as part of the pixel flat; properly, the fringing signal should be scaled to the exposure time and subtracted out (though for corrections around  $< 1\%$  or so the two reduce mathematically to the same thing). Igor offered an algorithm he had coded elsewhere, which would use filtering to isolate the wavy fringing signal from the rest of the

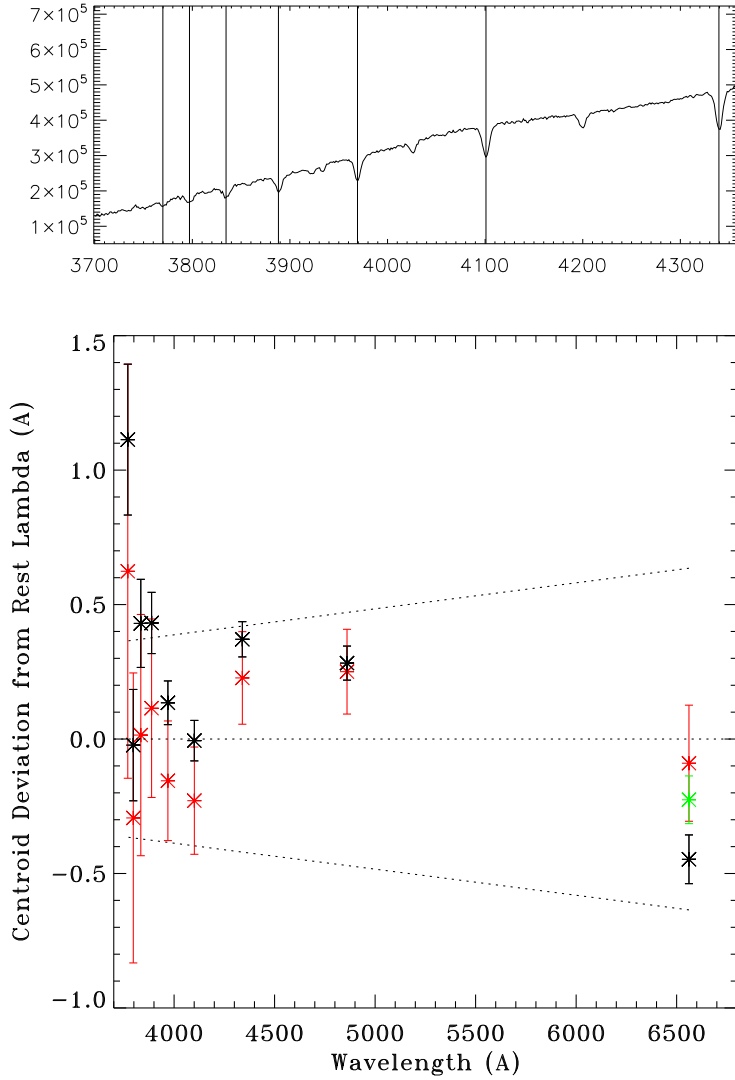


Figure 2: Top: Spectrum of HD192281 (reduced with HSRED 1.2) with the expected positions of Balmer absorption lines marked. Bottom: Difference between measured centroid of each Balmer line and expected position for HSRED 1.2 (black) and SPECROAD (red) reductions. Centroids were measured using a Gaussian fit with low-order continuum removal-green point shows H-alpha centroid measured using a slightly different fitting window; the large jump compared to the black H-alpha point suggests much of the variation in this plot is due to the crudeness of the centroid fit. Centroiding becomes particularly uncertain for the bluest, weakest Balmer lines.

flat field correction. I worked this into the pipeline, and verified that the fringing pattern it generates does indeed appear correct. Figure 3 shows a comparison of some red skyline residuals both with (left) and without (right) the fringing correction applied. Surprisingly, sky line residuals are marginally better *without* the correction applied, at least for typical exposure times. This result only became apparent recently, after we implemented the ‘skyfit’ fiber-throughput algorithm from SPECROAD (see below). Nevertheless, we will retain the fringing correction as an option, as long exposures on very faint targets could well see a benefit to using it.

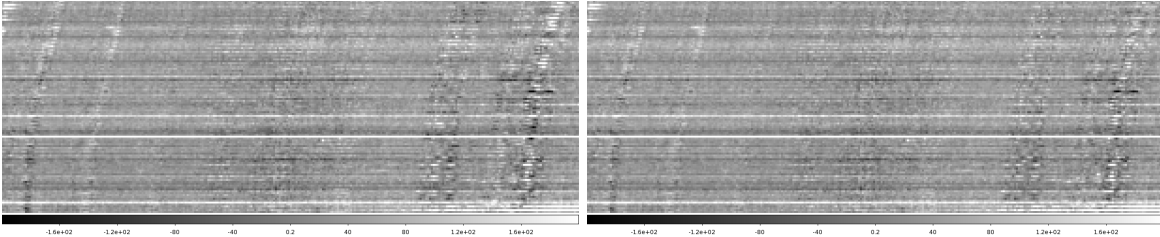


Figure 3: Red end of the spectra for an example config (wavelength on x-axis, fiber number running along y-axis), both with (Left) and without (Right) our red fringing correction. Sky line residuals are marginally better *without* the correction applied, at least for typical exposure times. Long exposures on very faint targets may see a benefit to using the fringing correction.

The main goal of obtaining sky flats is to measure the varying throughput of the Hectospec fibers. Correcting to a uniform flux scale is essential if we wish to use designated sky fibers to subtract the sky from target spectra. Due to variations in fiber positioning and conditions over the course of the night, however, sky flats do not provide a perfect correction. To further refine the relative flux scale, both SPECROAD and HSRED attempt to use the strengths of night-sky emission lines to make a correction to fiber throughput. The original IDL implementation of this was fairly crude, however, and gave sub-optimal results compared to SPECROAD. Therefore, we have implemented in IDL an algorithm believed to be very similar to the ‘skyfit’ algorithm from SPECROAD. For each fiber, our IDL SKYFIT routine compares the median sky spectrum (from all the sky fibers) with the object spectrum that has had a smoothed continuum subtracted (so that we are only comparing sharp lines to sharp lines). The correct scale factor ( $S$ ) is determined iteratively by minimizing the correlation ( $R^2$ ) between ( $S \cdot \text{fiber} - \text{median sky}$ ) and ( $\text{median sky}$ ). In other words, we find the point where the sky residuals in the sky-subtracted fiber are at their minimum. As in SPECROAD, this procedure appears to be quite robust even for bright objects, so long as some night sky emission lines are measurable. Typical corrections to fiber throughput are  $\sim 3\%$ , though a few fibers will often have a larger correction. We have mainly quantified the performance of SKYFIT in combination with other improvements to sky subtraction, and those metrics will be discussed in the following sections.

### 3 Cosmic Ray Removal

Another component of SPECROAD that we found to perform better than the original HSRED counterpart was the cosmic ray removal routine. HSRED adopted the LACosmic routine, which has the advantage of working on a single exposure, but its performance is quite slow. More importantly, we found it to have serious issues with bright objects, in particular standard star observations, where large chunks of the spectrum would be erroneously marked as CRs. To solve for these issues, Nelson Caldwell’s HCOSMIC routine had already been ported once into IRAF for use in SPECROAD. Consulting Nelson’s original routines, I did the same to produce an IDL version of HCOSMIC. HCOSMIC identifies CRs by calculating the statistics in a moving window, and marking CRs wherever the difference between the same pixel in two exposures is several times larger than expected. We have verified that HCOSMIC works as hoped on bright standard star spectra, and we have found no significant differences between the performance in IDL and

that in SPECROAD. In cases where only a single exposure is available, HCOSMIC cannot operate, and so we revert to using LACosmic (which, if desired, remains an option for any run of the pipeline).

## 4 Red Leak Modeling and Correction

The idea behind our new red leak correction is that the two red light sources are in largely fixed positions with respect to the focal plane, and so it should be possible to model/measure its intensity as a function of focal plane position, and thus predict the intensity observed through any given fiber using only fiber  $x$  and  $y$  coordinates, plus the exposure time. Using test data originally obtained in 2010 (with fibers in ‘ring’ configurations of various radii), we quickly determined that a simple analytic ( $r^{-2}$ ) parameterization of the red leak would not be sufficient.

However, by extracting and normalizing the red spectra through all the test data fibers, we were able to verify that the spectral shape is effectively constant over time and position, even between the two light sources. Thus, it should be sufficient to adopt a scalar red leak intensity, and map how this varies across the focal plane.

New test data (5min exposures against a dark dome) were obtained in Feb 2014, with the goal of sampling as densely as possible all portions of the focal plane. Red leak amplitudes were measured in each test fiber (a total of 3300 points), and mapped in `fiber_x` and `fiber_y`. These were transformed to `fiber_radius` and `fiber_azimuth` coordinates, and then a spline was fit in azimuth with a second simultaneous (polynomial) fit in the radius direction. The resulting focal plane map is shown in Figure 4, with azimuth on the horizontal axis and radius increasing from bottom to top. The two bright regions nearest to the light sources are evident, but one can also see a fairly complex structure to the light pattern.

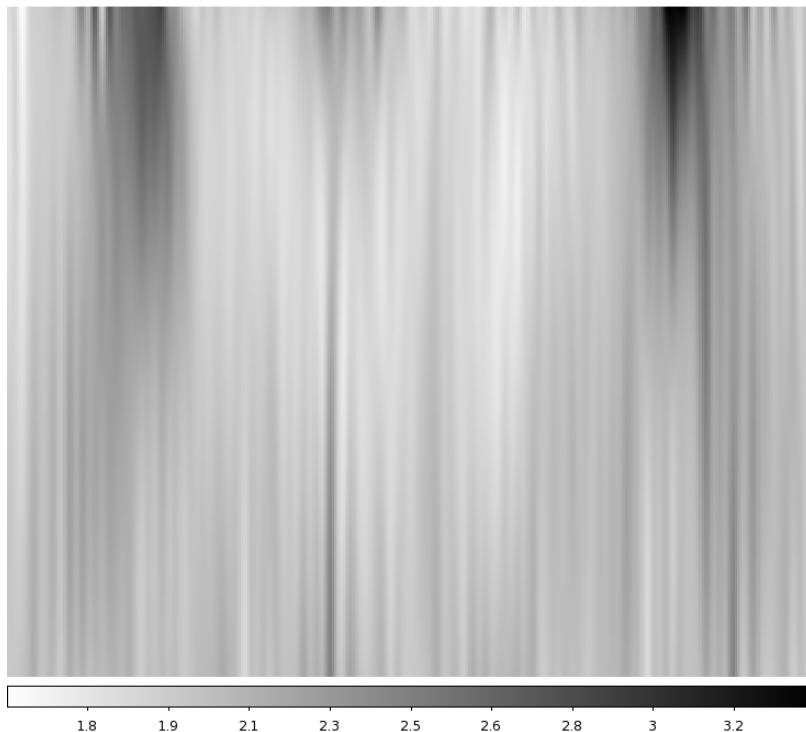


Figure 4: Model of red leak intensity as a function of azimuth in the focal plane (along the x-axis), and radius (y-axis, increasing from bottom). Color is proportional to the Log of red leak intensity in counts per 300s exposure (darker shades are brighter regions). Two bright spots can be seen at the outer edges of the focal plane, at  $\sim 60$  and  $\sim 300$  degrees azimuth. Other structures in the illumination pattern are also evident.

This complexity, unfortunately, has so far limited the accuracy of our corrections. The ‘sharpness’ of the model focal plane map can be varied by adjusting the number of breakpoints in the spline fit. Even though the map shown in Figure 4 has been tuned in sharpness to provide the greatest accuracy possible (given our available test data), jackknife tests show that the model only predicts the correct red-leak flux to  $\sim 20\%$ . This is worse performance than hoped for, and further digging has led to two findings:

During a normal run of the pipeline, we tweak fiber throughput corrections by measuring skyline amplitudes (using the SKYFIT algorithm discussed above); typical corrections are  $\sim 3\%$ , with a small number of fibers having a much larger correction. Without any on-sky flux to help us, the red leak test data cannot have this correction applied, and so this  $\sim 3\%$  (with outliers) represents the theoretical best we can do in terms of repeatable measurements of the red leak. We note also that, under full-moon conditions, even fibers that have been corrected for fiber throughput exhibit moonlight residuals that would be consistent with another, uncorrected  $\sim 3\%$  uncertainty in the fiber throughput (see below), and so we estimate that  $\sim 6\%$  may be the best we can hope for in terms of red leak accuracy.

Examining the residuals from our spline fit as a function of position in the focal plane, it is clear that certain areas, generally long, thin spikes in the radial direction, exhibit sharp variations in red leak intensity across very small changes in position in the focal plane. The precise sharpness of these features is hard to pin down, since our test data does not sample these regions with high enough spatial resolution. Nevertheless, we used this as motivation to ask for additional test data to help improve the spatial density of our test points. Unfortunately, problems with the instrument during the April 2014 run precluded any additional observations.

While overall the red leak is only corrected to 20%, there are many fibers in any configuration where this level of accuracy is good enough to provide an essentially complete correction. Using statistics from several configs that contained  $>200$  sky fibers, we find that  $\sim 75\%$  of the fibers show no evidence of residual red leak all the way to the end of the spectrum, another  $\sim 15\%$  show substantial but not ideal correction (generally, these show a residual only beyond 8900Å), and about 5% where the red leak estimate is completely off-base. We hope that these last 5% will be eliminated with improved test data, but this is to be determined. In Figure 5 (Left), we show an example fiber with a typical ‘complete’ correction. In the middle panel, the correction is not perfect, but this fiber had a particularly strong red leak, and so one can see that it is not just the strength of the red leak that determines how well we can make our correction: the right-hand panel shows a fiber with more modest red leak flux, but where the predicted correction is only about half what is needed.

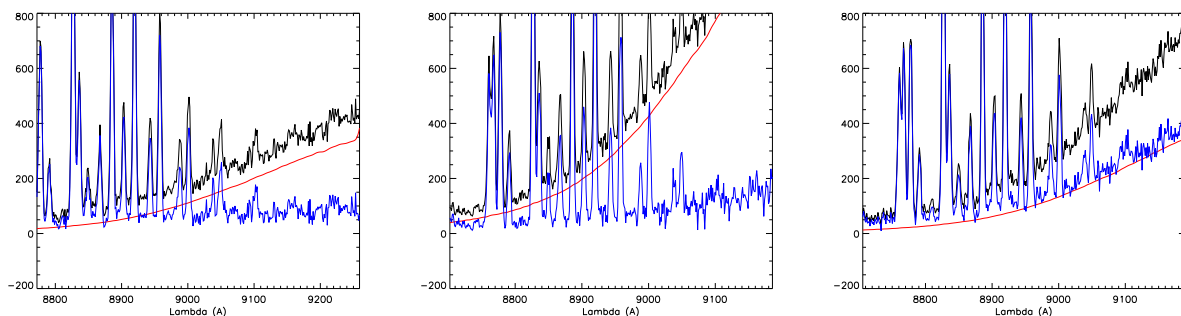


Figure 5: From left to right, examples of a good, fair, and poor red leak correction. The red end of the original spectrum is plotted in black (before sky subtraction), the red leak spectrum predicted by our model is shown in red, and the corrected spectrum is shown in blue.

## 5 Sky Subtraction

In SPECROAD, sky subtraction uses the six closest sky fibers to each target fiber (on the same chip), averages those skies, and subtracts it (after linearizing and rectifying all spectra). HSRED took a different approach, modeled on the SDSS method where all spectra are kept in their native, nonlinear wavelength mappings. This allows the construction of an oversampled ‘supersky’ model, where all the ( $\lambda$ , flux) pairs from all the sky fibers (again, one chip at a time), are fit with a spline function. This spline is then evaluated at the wavelengths for each target spectrum, and the resulting sky spectrum subtracted. The order and number of ‘breakpoints’ in the spline fit must be tuned, but in practice setting the number of breakpoints equal to the number of pixels in a single spectrum gives good results. At each pixel/wavelength increment (1.2Å for 270-line data), this choice results in a single 4<sup>th</sup> order spline fit to N\_SKIES oversampled data points.

Both techniques for sky subtraction work well, but have their drawbacks: since it combines *all* the skies to make its model, rather than just the nearest few, the HSRED method in general left larger sky-line residuals. In contrast, SPECROAD required all spectra to be linearized before sky subtraction, resulting in less precision in the final spectra.

In HSRED 1.2, after quite a bit of experimentation, we have made one key improvement to the sky subtraction algorithm that results in performance better than either predecessor. Inspired by Igor Chilingarian’s MMIRS reduction pipeline, we adopt the HSRED spline fit, but add a second dimension in fiber number. Thus, the sky at each point is now a smooth, oversampled function of both wavelength and position along the chip. This enables much better performance in terms of sky line residuals, but the cost is more parameters in the fit: we have moved from a 4<sup>th</sup> order spline to a third-order spline plus a second-order polynomial fit in fiber number. The number of sky fibers required for a reliable fit thus increases; for configurations with too few good skies, we revert back to a single-dimension spline fit (as before). We also revert to a single spline (sometimes only for part of the chip), in a few other edge cases—such as when all the sky fibers are stacked towards a single end of a chip.

The choice to use fiber number as the second dimension in the fit was not an obvious one. We tried a number of other parameters, including dispersion/resolution measured from the arc line widths, but fiber number clearly provided the best results: when examining data reduced with the old, single-spline fit, it is quite obvious that sky line residuals vary smoothly across the chip, though it is unclear why this should be so.

In any case, the current method provides superior results. In Figure 6, we show a scatter plot of (residual) S/N for all sky-subtracted skies from a single bright-time configuration (270-line grating) on 2014/02/19. Contours show  $\pm 1$ - and  $\pm 2$ - sigma for the distribution of points. The dominant component in the noise spectrum is the Poisson noise from the bright moonlight background, and so the fact that the  $\pm 1$ -sigma contours lie very close to the  $S/N = \pm 1$  lines indicate near-Poisson limited subtraction. However, since the sky model is built from these same sky spectra, the sky-subtracted sky is not a completely independent estimate of residual noise in the actual target spectra. In the bottom panel, we show a similar S/N scatter plot for the object spectra from the same Hectospec configuration. Though these exposures collected very little flux from the target spectra, due to the bright moon and significant clouds that evening, the upper envelope of points in this scatter plot is clearly dominated by the object flux that was obtained. Instead, it is instructive to look at the lower envelope of points, in particular the lowest,  $-2$ -sigma contour line. Here, one can see that there is a residual, long-baseline systematic, manifested by the lower envelope of points rising from  $S/N \sim -2$  to  $S/N \sim 0$  in the 5000-7000Å range. This appears to be excess, unsubtracted moonlight that is not well-predicted by our sky model. (Note that despite appearances in Figure 6, the magnitude of the error varies from fiber to fiber, and so a simple fixed correction is not feasible). We note that dark-sky nights do not show this problem. Also keep in mind that this example observation with both full moon and clouds essentially represents a worst-case scenario for sky-subtraction performance.

Two avenues were explored to explain and/or better remove this systematic. The first was improving the accuracy of fiber throughput corrections so that the sky-fibers can better predict the amount of moonlight

that must be subtracted from the target fibers. (This was our motivation for implementing the SKYFIT algorithm in IDL, as discussed above, and that effort reduced our systematic bright-time error roughly in half, resulting in what you see in Figure 6.) The second possibility is that the residual is due to scattered light off of the dome or other source, such that the flux incident on a fiber does not vary smoothly across the field of view or as a function of fiber number. In either case, no further improvement seems to be achievable at this time.

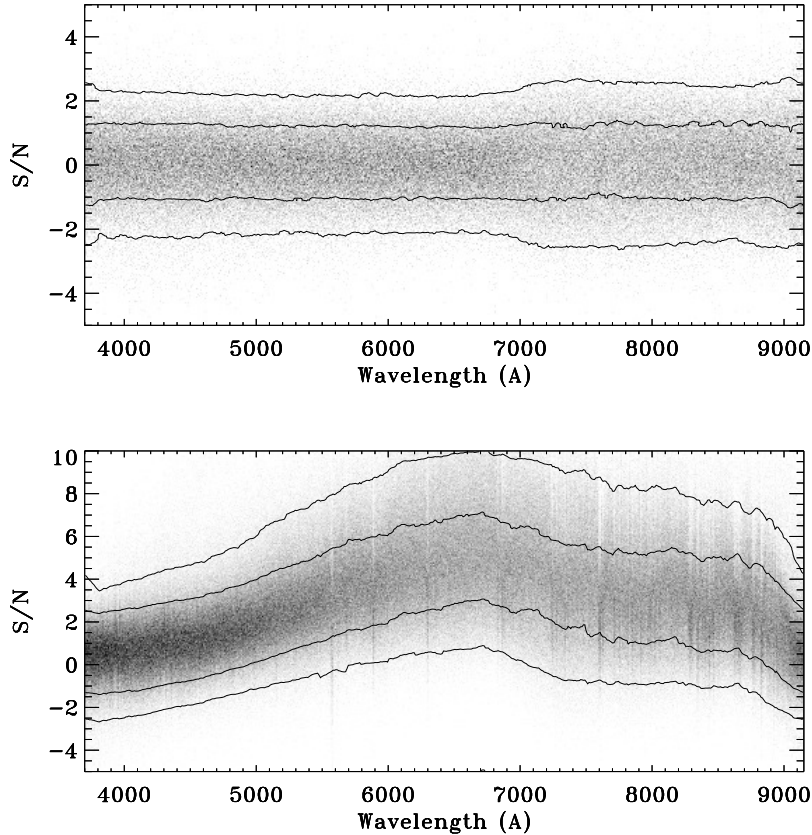


Figure 6: Top: Scatter plot of S/N for all sky-subtracted skies from a single bright-time configuration (270-line grating) on 2014/02/19. Contours show  $\pm 1$ - and  $\pm 2$ - sigma for the distribution of points. The residuals here are very close to that expected from a Poisson noise distribution. Bottom: Similar S/N scatter plot for the object spectra from the same Hectospec configuration. Note that the lower,  $-2$ -sigma contour is not flat, but shows a systematic residual, especially in the 5000-7000 Å range.

## 6 Telluric Absorption Correction

A new feature not in SPECROAD or the original HSRED is a correction for the telluric absorption found in the A-band and B-band regions of the spectrum. We apply a very simple method using a single hot-star template spectrum to derive a normalized absorption profile (by dividing the stellar spectrum by a smoothed fit to itself). This is then scaled according to ‘Beer’s Law’, which depends only on the airmass of the observation (and that of the template spectrum observation). The resulting absorption function is then applied to the target spectra. The basic method is described for Keck HIRES data at: <http://www2.keck.hawaii.edu/inst/common/makeewww/Atmosphere/index.html>. It is also similar to the iraf ‘telluric’ task, but with the important difference that the template spectrum is obtained with the same instrument and matched in resolution.



While in principle the telluric correction varies with atmospheric conditions like temperature, pressure, wind speed, etc, in practice for these strong, unresolved bands in the optical, the very simple model provides surprisingly good results. We show in Figure 7 a comparison between data reduced with and without the telluric correction. Even across many fibers, there are no obvious residuals or mis-corrections using this method.

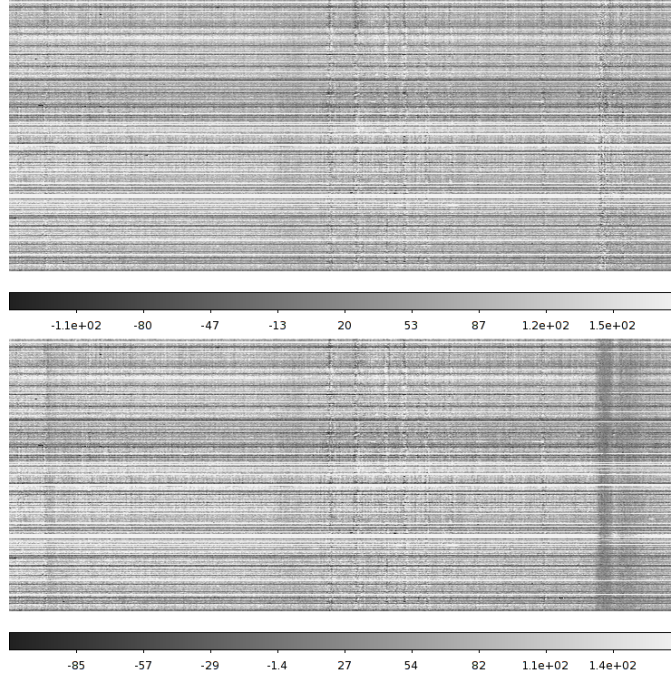


Figure 7: Reduced spectra from an example configuration (wavelength on x-axis, fiber number running along y-axis), both with (Top) and without (Bottom) our telluric correction. In the lower panel, the location of the A-band absorption is clearly visible on the far right, and the somewhat weaker B-band is toward the far left.

Matching the spectral resolution of the template and the observation seems to be the most important factor for an accurate correction; we save, within the pipeline tree, a single template spectrum taken with the 270-line grating. Thus, only 270-line data is currently corrected, and this part of the pipeline must be skipped over during 600-line reductions. It would be possible, however, to implement this correction on 600-line data in the future.

As described in Stevenson (1994), correction using this method is subject to some inaccuracy because each telluric ‘band’ is really an unresolved forest of lines. If a very intrinsically narrow emission line ( $< 30\text{km/s}$  or so) falls into the telluric region—even though both the emission line and telluric absorption are smoothed to lower resolution by the instrumental response—that emission line is really superimposed somewhere on the forest of lines, and it could have fallen on, or in-between any of the narrow telluric features. At low resolution, there is no real way to know what the true flux in that narrow emission feature was. Though in practice there is nothing to be done without more complex modeling, it is worth being aware of this issue, and exercise caution in interpreting narrow line measurements from the telluric bands, despite the correction we have applied.

## 7 Coaddition

While SPECROAD coadds multiple exposures by simply summing counts (after the spectra have been linearized), the original HSRED scheme was much more complicated. Due to the design choice of keeping spectra in their native wavelength mapping, simple summing of exposures would be problematic. Instead,

they decided to perform an unusual coadd trick similar in concept to the sky-subtraction: an oversampled spline model is fit to all the data points from all exposures, and that model is then evaluated at a set of wavelengths. (The choice of output wavelength grid could have been linear, but instead was chosen to be the wavelength array for the first exposure). In order for the oversampled spline to make sense, however, all the different exposures must be scaled to have the same overall flux scale (so that the spline interpolation between data points gives sensible results.)

Another feature included with HSRED was the ability to flux calibrate spectra based on magnitudes provided in a photometric catalog. When a catalog like this is available, scaling the exposures to the same flux level is not difficult, and co-adding goes well. However, for most Hectospec observations, no such catalog is available. In these cases, it was necessary to make a guess at the scaling between different exposures, via a crude estimate of S/N, and then attempt to scale all the exposures. Often, this procedure works fine, but we found that in the case of variable or poor observing conditions, the scale factors applied could be wildly incorrect.

Thus, we decided to replace the original IDL algorithm with a simpler method. For each set of exposures, we attempt to pick the one in the middle, and assign it to be the reference wavelength grid. We then, one fiber at a time, interpolate the other exposures onto the reference wavelength grid. Typical shifts are very small, and so this introduces only a small uncertainty/degradation of resolution. We then calculate both inverse-variance-weighted sums (primary product) and directly summed counts (secondary product), as well as new variance spectra. With this method, there is no need to match the flux scales of the various exposures, and the resulting co-added spectra accurately reflect the total number of photons reaching the detector, regardless of the observing conditions. Though the final spectra are on an irregular wavelength grid, we note that these coadds are not strictly in native pixel coordinates anymore. If true native pixel mapping is important, however, the individual exposures are saved and available.

All together, the pipeline changes described above result in final, reduced spectra of improved quality compared to either predecessor, and data reduction appears to be robust across changes in observing conditions and time: data has been successfully reduced going all the way back to 2004, and efforts will soon be underway to reprocess all data in our archives (See below.) In Figure 8, we show the full set of reduced target spectra for a single recent observation. These spectra have been rectified in wavelength and stacked, such that fiber number runs along the y-axis, and wavelength runs from bluest at the left of the top panel to reddest at the right of the bottom panel.

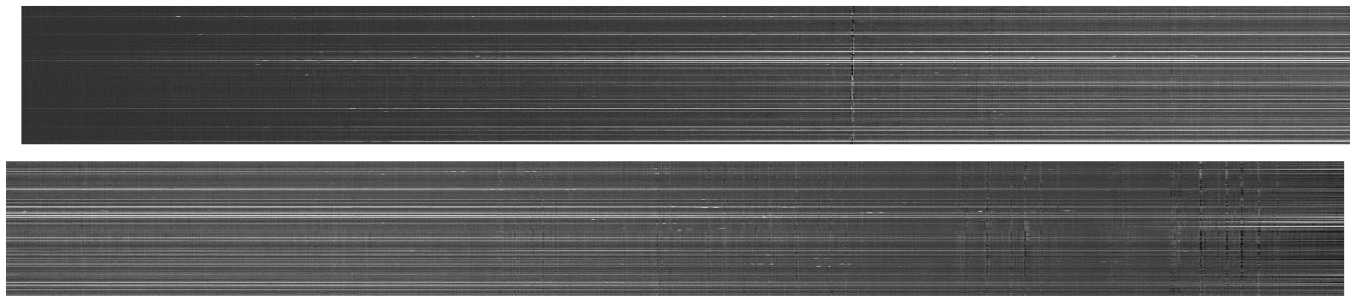


Figure 8: 2D image showing 259 target spectra from arrayed along the Y-axis, rectified in wavelength running along the X-axis (blue to the left). Data is from an arbitrary ‘hectomap’ field taken during dark time. Top panel shows the blue half of the image, with wavelengths running roughly from 3700Å to 6700Å, and the bottom panel shows the red half (slightly overlapping), with wavelengths from 6100Å to 9150Å.

Statistically comparing to the identical data reduced with SPECROAD, we find that the residuals are close to normally distributed. In Figure 9, we plot a histogram of the pixel-by-pixel differences between the new HSRED and SPECROAD, normalized by the reported uncertainty in each pixel. The Gaussian fit over plotted has  $\sigma$  nearly unity, which indicates that overall the two reductions match very well. Toward the lower right tail of the histogram, one can see that SPECROAD shows an excess of points with high flux

compared to HSRED. These are due to 1) better skyline subtraction in HSRED, and 2) a few fibers with differences in the fiber throughput correction/normalization. The algorithm for scaling fibers is intended to be the same, so the source of those deviations is unclear (though these differences may be mainly in ‘unused’ fibers).

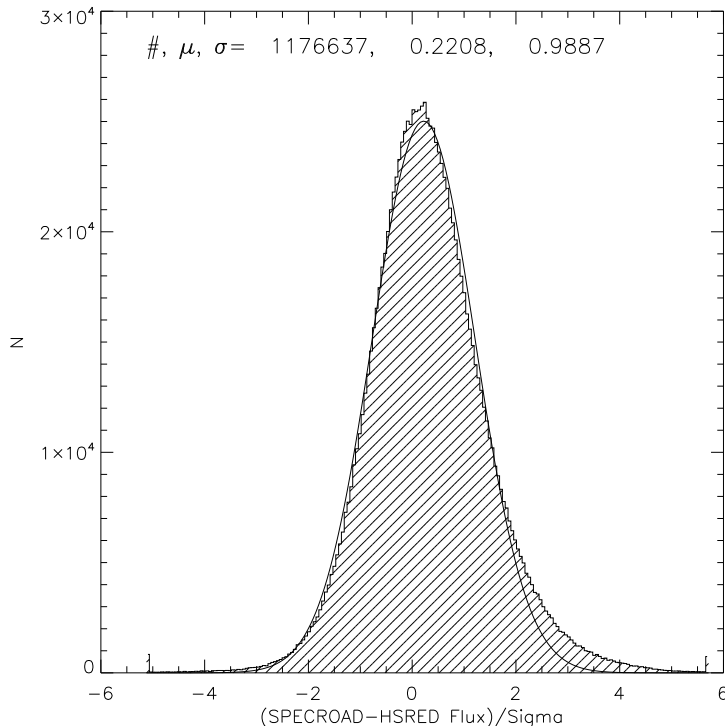


Figure 9: Histogram showing the difference in flux between SPECROAD and HSRED reductions, for all the pixels in a full set of 300 reduced fibers, and divided by the quoted noise in that pixel. The Gaussian fit over plotted has mean ( $\mu$ ) and  $\sigma$  as annotated at the top of the plot.

## 8 Data Products

The native output format for HSRED is a series of multi-extension fits files containing the reduced spectra, inverse variances, fiber configuration info, and various other supplementary products. Wavelengths, as mentioned above, are non-linear, and are stored directly as pixel values in the zeroth fits extension. These products are similar to those from other reduction pipelines based on the SDSS code, and are likely familiar to many researchers used to working in IDL. However, this format is not at all practical for those who wish to work in IRAF, and it is quite different from the SPECROAD data products that we have been distributing for years.

Thus, for maximum flexibility, we have greatly expanded on a routine called ‘HS\_TOIRAF’ from the original HSRED code. We have been able to essentially reconstruct the data in an almost-identical format to that produced by SPECROAD, with some improvements. We provide individual fits files for each object spectrum (linearized) in IRAF ‘multispec’ format, but we now include the noise spectrum for each object, which was not available before. Likewise, we provide a single combined ‘multispec’ file for all the fibers, again with the noise spectra included. The equivalent file from SPECROAD was produced *before* sky-subtraction; the new file is generated from the final reduced products (it includes the sky spectra, so the old-style files can be recreated with ease if needed.)

All of these products are described in a new README file, which we distribute with every data set.

We note, also, that we have modified the HSRED code to add fits header keywords to indicate which reduction steps have been completed, and in some cases, under which mode (e.g., for sky subtraction, we record the number of sky fibers used, and whether the fit was 2D or reverted to 1D.) This provides a concise record of the reduction, and is similar to SPECROAD.

## 9 Comparison of 1D pipelines and redshift finders

One of the key data products we have distributed with our SPECROAD data products has been redshifts/radial velocities computed with the RVSAO IRAF package. These redshifts were vetted by eye by Susan Tokarz, and represent one of the greatest legacy products of Hectospec. Since we have been able to transform data products from the new HSRED pipeline into IRAF multispec format, we can continue to run RVSAO/XCSAO on all the data we process at TDC. Note that RVSAO, being an iraf task, is not very portable, and so this aspect of the pipeline will not be part of the code we distribute publicly.

Being able to run the identical cross-correlation code on products reduced with two different pipelines has turned out to be an excellent diagnostic tool to ensure that the new code is performing as expected. In Figure 10, we show a histogram of velocity differences (in km/s) between spectra reduced with SPECROAD and the same spectra reduced with HSRED (both run through RVSAO). We have limited this comparison to those with decently reliable velocities from the cross-corelation ( $R > 5.5$ ), in order to filter out junk spectra. The results show that the two pipelines produce redshifts that are statistically identical, with a spread of  $\sim 40\text{km/s}$ , comparable to the quoted error on a single radial velocity estimate.

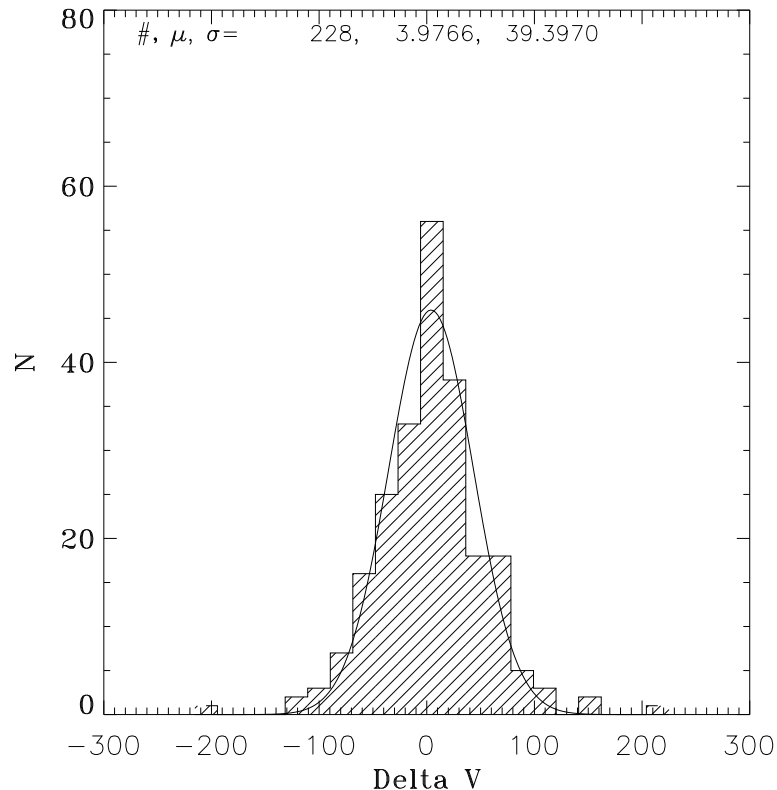


Figure 10: Histogram of difference in radial velocities, SPECROAD-HSRED, for all good spectra ( $R > 5.5$ ) in a single example configuration. At the top of the plot are the parameters of the Gaussian fit (smooth curve), indicating an offset of only 4km/s and a spread of  $\pm 40\text{km/s}$  (1-sigma). The median quoted uncertainty from XCSAO is  $\sim 40\text{ km/s}$ , so the differences observed are consistent with expectations.

We must note that a wavelength/velocity correction to the heliocentric reference frame is applied to each exposure in HSRED 1.2, after the telluric correction (which should be applied in the Earth’s reference frame), but before co-addition. The earlier version of HSRED had the heliocentric correction placed too early in the pipeline, such that the correction was often erased at the point where wavelengths were tweaked to match sky line positions. It is also worth pointing out one key difference between the heliocentric correction in SPECROAD and that in HSRED 1.2. In SPECROAD, a heliocentric correction is made to the velocities/redshifts identified by the RVSAO routine, but no correction is made to the data itself. In HSRED, in contrast, the *data itself* is corrected to the heliocentric rest frame. When this data is then run through RVSAO locally, we *turn off* the RVSAO heliocentric correction, to avoid applying a double correction.

A more detailed comparison between redshifts from the new vs old pipeline will be performed in the future. Mike Kurtz is interested in digging into the detailed statistics of, for example, how often the best-fit template spectrum differs between the two reduced spectra, and how often the code misses a redshift that would have been found by Susan Tokarz. I have run all of the data from trimester 2013C through the new pipeline, to provide a starting data set for this comparison, but this task is still waiting on implementation of our new reduced-data POSTGRES database. As part of this, we will also examine the potential usefulness of the 1D analysis pipeline included with HSRED, and based again on SDSS. While this SPEC1D code appears to be quite sophisticated, matching against a larger number of templates than RVSAO, in order to estimate, for example, the spectral type of stars, or to perform galaxy/AGN classification, the code is very time-consuming to run (on the order of hours per Hectospec configuration), and so it is unclear if we want to begin running it routinely.

## 10 Operations at TDC and Distribution of Code

HSRED v1.2 is currently the operational pipeline at TDC for all Hectospec data reductions. This has been the case since Feb 2014, though minor bug fixes in the time since then may mean that data from Feb/Mar should be reprocessed. In the near future, we hope to evaluate whether HSRED can also be adapted for use with Hectochelle data. In anticipation of this, we have already written code capable of transforming non-linearized HSRED spectra into a (still nonlinear) format readable by IRAF.

A typical run of the pipeline completes in only about an hour or less, and multiple fiber configurations from the same night can be reduced in parallel. The code is completely automated, requiring only that bad/unwanted exposures (as well as skycam images) be deleted from the processing directory before starting the pipeline. Exposures obtained with different gratings/central wavelengths must also be sorted into separate directories before processing. For each night, we copy all data from our raw data archive into a working directory, consult the commlog.ps file to delete/sort files, and then execute the pipeline. Each time the pipeline is run, a ‘run number’ must be specified, such that data can be easily reduced multiple times with different pipeline parameters (for example, enabling/disabling red fringing correction). After reduction, our new, expanded set of data products can be distributed to PIs using a slightly modified version of our usual distribution scripts.

Our aim for the code base is to transfer it back to Richard Cool at MMT, for hosting and public distribution directly from MMT. Before this code transfer, some work remains to clean up the code comments/documentation. The original HSRED relies on very old versions of two external packages maintained by SDSS; I hope to test whether the code might be made to work with the most recent version of these packages. (There is currently an incompatibility that I have yet to track down).

Finally, I intend to write a short, updated cookbook for observers hoping to use the new pipeline. Wrapper scripts I have written should make running it considerably simpler than the old version of the code, but changes to the procedure and the data products will have to be summarized for those who are accustomed to the old version.

As mentioned above, we hope to reprocess the entire Hectospec archive with HSRED 1.2, regardless of the PI’s institution. This will produce a complete, uniform catalog of all Hectospec observations spanning

over a decade. Over time, we hope to make this entire set of spectra publicly available, and I expect one of our key projects going forward will be to complete the underlying database work and connections to VO protocols/services that would enable us to distribute this data. As a first step, we hope to choose and release, with permission from the PI, a single ‘showcase’ dataset that will hopefully prompt broad community interest.