NRES Data Reduction Pipeline: Brief Overview
Tim Brown – Aug 2017

Summary

This document is intended to help new users of the Network of Robotic Echelle Spectrographs (NRES) to understand what operations are performed by the v1 data reduction pipeline, mostly so that they may understand what is in the archived data products, and how to interpret them. I will also call out sections of the pipeline that are currently problematic, hence where user caution is wise.

Instrument Architecture

A full NRES installation involves two 1m telescopes, each with an Acquisition and Guiding Unit (AGU) at the focus. Each AGU couples starlight from the telescope into a "star" optical fiber and thence to the spectrograph. Both star fibers are connected to the spectrograph at all times, so either telescope can be used to feed the spectrograph. In the future we hope to allow both telescopes to be used simultaneously, but that is not possible at present. Each AGU also contains an autoguider CCD camera to acquire the desired target star and keep its image on the star fiber, and internal switching optics that can couple calibration light into the star fiber, and allow special daytime transmission calibrations.

About 15m from the telescope domes is a temperature-controlled room (the "Igloo"), which holds the bulk of the spectrograph systems. These consist of the Calibration System, several layers of temperature control, a constant-pressure vessel, the Exposure Meter, and the Spectrograph.

The spectrograph is always coupled to the (one or) two star fibers, and also to a single "reference" fiber. Light from these 3 fibers appears in the spectrograph image plane as 3 parallel cross-dispersed echelle spectra, with the spectra from different fibers separated from each other by about 20 pixels (300 micron) in the image plane. The reference fiber always carries light from the Calibration System straight to the spectrograph. Each of the two star fibers may carry starlight, or it may carry reference light which originates in the Calibration System, passes through optical fibers out to the AGUs, and then is coupled into the corresponding star fiber in place of light from the telescope.

In a 1-telescope system, one of the three star fibers is never illuminated.

The Calibration System is an optical beam switcher that allows light from any of three sources (a tungsten-halogen lamp "TH", and two ThAr lamps "slave" and "master") to be coupled to three different outputs. One of the outputs goes to the reference fiber, and the other two go to the two AGUs, where they may be coupled

back to the spectrograph through star fibers, or not. Each Calibration System output has an independent shutter, so the three outputs may be passed or blocked in any combination. The reference fiber output has an adjustable neutral-density filter, so that we may reduce its intensity by up to a factor of 100, to prevent overexposure of the ThAr reference lines during long stellar integrations.

Light passes through several thermal control layers and the constant- pressure vessel wall via a trifurcated optical fiber bundle. It is admitted to the spectrograph proper through through the main shutter, which blocks or passes the light from all 3 fibers together.

Parasitic light reflected from the AR-coated first surface of the cross-dispersing prism is reflected back into the Exposure Meter. This unit re-images the three spectrograph input fibers onto another trifurcated fiber bundle, which carries the light out of the pressure vessel, where the fiber tips are imaged onto a CCD camera. In this way we obtain a continuous record of the light intensity emerging from each of the 3 input fibers.

The spectrograph's main CCD detector is a 4K x 4K Fairchild chip with 15 μ square pixels. We cool it to -90C with a cryotiger. Dedicated thermal servos hold the coolant lines and the cryostat housing at the same temperature as the pressure vessel interior.

The main CCD controller, the thermal and pressure control electronics, the cryotiger, and the Calibration System all reside inside the 2nd thermal control layer (the Igloo wall is the 1st), where the temperature is stable to about 0.1 C RMS.

A complete NRES installation therefore involves 2 telescopes, 4 CCD cameras, 3 light sources, and a collection of shutters, drive motors, and temperature-, pressure-, and position-sensors. This complexity is reflected in the reduction pipeline and the resulting data products described below.

Data Types

The data products used and produced by the pipeline are distinguished first by their "type", which is encoded in the first few (upper case) characters of their file names, and sometimes in their FITS headers.

The NRES spectrograph and its associated Calibration System and site software produce four kinds of calibration image (BIAS, DARK, LAMPFLAT, DOUBLE) and one kind of science image (TARGET).

BIAS and DARK are what they say.

LAMPFLATs are images in which two input fibers (the reference fiber and one or the other star fiber) are illuminated by the TH lamp.

DOUBLEs are images in which two input fibers (the reference fiber and one or the other star fiber) are illuminated by a ThAr lamp. The lamp may be either the slave or the master lamp.

TARGETs are images in which the reference fiber is illuminated by a ThAr lamp (either slave or master), and one of the star fibers (in the future, possibly both) by starlight from a telescope.

In the parlance of the LCO site software, each of these 5 data types has a corresponding molecule – a piece of code that causes the system to produce an image of the desired type.

The four calibration data types are stored as simple FITS files. TARGET data are more complicated because they contain additional data from telescopes, autoguiders, and the Exposure Meter. TARGET files are therefore stored as multi-extension FITS files, using binary FITS tables.

In addition to the basic data types that come from the spectrograph, the pipeline creates and uses a number of derived data types: TRACE, FLAT, TRIPLE, EXTR, BLAZ, SPEC, THAR, ZERO, and RADV. These have various formats, and are described below.

Data Flow

Raw data come from spectrographs at site in near real time. Each new data file arrival triggers the pipeline at LCO headquarters in Goleta. For new calibration files, the pipeline copies the incoming data into local disk storage, and saves relevant metadata in a searchable database for later use. New TARGET data are also processed immediately, using the best then-available calibration data, based on a suitable database search. Some intermediate data products (described below) for each TARGET file are saved to local disk storage, and their metadata also go into the database. At the end of TARGET processing, the pipeline bundles the raw data, selected extraction-level data products, and summary data products into three compressed tar files, which are sent to the LCO Archive.

Separate processes run on a daily basis to construct the various needed master-calibration files and other derived calibration data. The inputs for these processes are found via suitable searches of the database, and their products are placed in local disk storage, entered in the database, and written to the Archive in case they are needed by users.

There is a facility to rerun the pipeline analysis of old data, as seems necessary or desirable.

Ingestion

Data ingestion occurs whenever a new data file of standard type (BIAS, DARK, LAMPFLAT, DOUBLE, or TARGET) from any spectrograph is recognized in the input data buffer area. The pipeline copies the main data segment into local disk storage, extracts metadata that will be used in the pipeline analysis, and writes relevant metadata describing the file into the database.

Super-Calibration Files

Calibration data files are normally acquired in bunches of the same type, during daylight hours. Special (not pipeline, per se) routines run on a daily basis to average these into super-BIAS, super-DARK, and super-FLAT files.

Super-BIASes and super-DARKs are in the same format as BIAS and DARK files, but are distinguished from their un-averaged progenitors by their metadata and corresponding database entries.

super-FLATS are combinations of extracted spectra involving fibers (0,1) and (1,2), so that all three fibers have valid flats (except, of course, for 1-telescope installations).

TRACE files describe the positions and cross-dispersion shapes of the spectrum orders on the CCD. They are constructed from super-FLATs.

TRIPLE files describe the along-dispersion offsets between ThAr spectra for fibers (0,1,2). They are constructed from multiple DOUBLE files, involving both fibers (0,1) and (1,2).

ZERO files are used in estimating radial velocities of stars; they contain a version of the standard stellar spectrum that will be compared to observed spectra that are derived from TARGET files.

Calibration and Extraction

For input TARGET data, the pipeline locates the "best" BIAS, DARK, FLAT, and TRACE files from a database search, according to simple rules. Using these calibration data it (a) Corrects for bias and dark in the obvious way. (b) Using the TRACE data to know where the orders are supposed to be, it fits and subtracts a model of the between-order background light. **The numerical procedure used here is very crude, and could use improvement.** (c) Using the TRACE data to define "extraction boxes" that are supposed to be centered on the order positions, it performs a preliminary extraction and order cross-dispersion centroid computation. If the computed centroid displacements are too large, it applies a parametric adjustment of the TRACE data (which moves the boxes) and iterates. (d) When satisfied with the extraction box positions, the pipeline computes an optimally-extracted spectrum (dimensioned nx by norder by nfiber), i.e. typically (4096 x 67 x 2) (because one fiber is dark), using cross-dispersion profiles from the TRACE file, and a noise model derived after step (b) above. (e) The pipeline examines the residuals around the extracted spectrum (d), looking for evidence of radiation events. Any that are found have the fitting weights for nearby data points set to zero, and the fit is re-computed. **This part of the analysis is working poorly, and needs work.** (f) Three versions of the resulting extracted spectrum are then saved to local disk storage. These are:

EXTR – This is a raw extracted spectrum; each order looks more or less like

the blaze function of the spectrograph.

BLAZ – This is like EXTR, but with a constant multiple of the FLAT spectrum subtracted from it. This spectrum should have a near-zero mean value, and goes to zero at the edges of the blaze function. It has desirable noise properties for use in radial velocity estimation.

SPEC – This is the ratio of EXTR and FLAT, ie something like the true stellar spectrum with the instrumental response removed. It is of course noisy at the edges, and prone to systematics arising from differences between star and flat illumination of the spectrograph optics.

Wavelength Solution

The pipeline adjusts parameters in a model of vacuum wavelength vs (x-coordinate, order index, fiber index) to give an optimum match between the positions of emission lines observed in the fiber 1 (reference) spectrum and those implied by the wavelength model and the ThAr line catalog by Redman (2013). This adjusted model defines the wavelength solution for fiber 1. Using the chosen TRIPLE file, this model is extrapolated to fibers 0 and 2. The pipeline saves the entire wavelength solution – an array dimensioned (nx x nord x nfiber), along with all of the model parameters, to a file of type THAR in local disk storage.

Exposure Meter

The Exposure Meter runs continuously, with light fluxes from each of the three fibers being measured and saved (along with a time tag and error flag) for every ExpM exposure. Site software collects these values for the time during which each TARGET molecule is running (which includes the entire integration time plus some overhead at each end), and places them in the "Exposure Meter" data segment of the TARGET multi-extension FITS file. The pipeline uses these data to compute the flux-weighted mean time corresponding to each of the star fibers. These values are passed to the routine for estimating radial velocities, and they are stored as metadata for the RADV output file written by that routine (see below).

Autoguider Performance

Site software takes output from each AGU and produces a time series of autoguider corrections applied during the execution of each TARGET molecule. These time series and their corresponding time tags are written into the two "AGU" data segments of each TARGET multi-extension FITS file. As of v1, nothing is done with these data. Plans call for various statistics describing these time series to be produced and stored in the header of the RADV output file.

Radial Velocity Estimation

In v1, radial velocity estimation is very much a work in progress. Accordingly, it is possible to run the RV estimation after the rest of the pipeline has executed, using as input the archived outputs of the pipeline up to this point.

Estimating a stellar radial velocity relies on comparison of the BLAZ output spectrum from a TARGET file with the spectrum from a ZERO calibration file. For consistency, it is important that a given TARGET star should always be compared with the same ZERO star. This is accomplished by linking (in the database) each target star with a particular ZERO file. The spectra in ZERO files are kept in the BLAZ format. For v1, these ZERO spectra are made from averages of spectra of an observed star, chosen to be of similar spectral type as the target star. In the future, we will construct a grid of ZERO files based on model stellar atmospheres.

Given a ZERO spectrum and a BLAZ spectrum (and its corresponding THAR file, containing the BLAZ wavelength solution), the pipeline first estimates an approximate redshift by cross-correlating the BLAZ and ZERO spectra, but only for the echelle order containing the Mg b lines (roughly 516 nm). Based on this preliminary estimate, the pipeline then interpolates the entire ZERO spectrum to the provisional redshifted wavelength scale, and breaks each order into a number of "blocks", ie contiguous wavelength segments. The pipeline then performs a fit to estimate the residual redshift of each block, and formal errors. Last, the pipeline constructs several estimates of the "mean" redshift, taking differently-weighted averages or medians of the individual block redshifts.

Outputs of the radial velocity analysis are written to a file of type RADV, which is a FITS extension file with an empty main data segment. The first extension table contains the cross-correlation function and various cross-correlation-related statistics. The second extension table contains the computed residual redshifts per order and block, and useful statistics related to them.

Spectrum Classification

In v1 we have no facility for classifying stellar spectra (ie, estimating Teff, log(g), log(Z), vsini. This capability will be added later.

Diagnostic Plots

At the end of each TARGET reduction, the pipeline creates two diagnostic plot files: The PLOT file is intended as a close simulacrum of Dave Latham's "quick" plot. It aims mostly to show various aspects of the target star spectrum, although one may infer some kinds of pipeline errors from it. The PLQC file contains mostly diagnostics of the accuracy of the TRACE file used for extracting 1-dimensional spectra, and of the wavelength solution.

Data Products

At the end of reducing a TARGET file, the pipeline bundles various output data into 3 compressed tarballs, and writes them to the LCO archive. They contain: (a) The raw input data file. (b) Various extracted data products, and calibration files that went into their construction. These are the files FLAT, TRACE, EXTR, BLAZ, TRIPLE, THAR, ZERO, and RVDAT. (c) The summary statistics file SUMSTAT, and the diagnostic plot files PLOT and PLQC.

Thus, if you want only the key results of the observation, and are optimistic enough to accept the pipeline results without checking, then you need only retrieve part (c) from the archive. If you want to start from extracted spectra and the pipeline wavelength solution, then get part (b). If you want to start from the beginning, get part (a).

If the pipeline is set up to bypass the radial velocity estimation, then part (c) will not be written to the archive, and part (b) will be missing the RADV file.