HRS Pipeline High-Level description

DOC NO: xxxxx Issue : B

Prepared by : HRS Pipeline Team

Approved by :

Version History

Author	Issue	Date	Change history
É. Depagne	A	11/06/2015	Creation
É.Depagne	B	21/10/2015	Update after Comments and discussion with Petri and Steve C.

Acronyms and abbreviations

HRS	High Resolution Spectrograph
HPS	HRS Pipeline Specification

Definitions

Pipeline	A Pipeline is a suite of software intended for producing science grade 1 dimensional wavelength calibrated spectra

Vous pouvez ajouter des titres (Format > Styles de paragraphe) qui apparaîtront dans votre table des matières.

1. Introduction

This document describes the HRS Pipeline. It presents the high-level requirements, as well as some details about the internal workings of the pipeline. It serves also as a description of the current state of the pipeline. It contains a tentative detailed man-power description of the work required to have a fully functional pipeline.

2. High level description

The HRS pipeline is a suite of software whose purpose is to produce science ready data from a set of raw science frames taken at the SALT telescope, and a series of calibration frames.

The language of choice for the pipeline to be written in is Python. This is the most used high-level language in the scientific community, and especially in astronomy. It allows easy interfacing with other languages, if needed, and provides many bricks as modules from which complex tasks can be built. And it is the language of choice for the already existing reduction suite for other SALT instruments. As every software written at SALT, it will be open-source.

It works in steps, each of which serves the purpose of preparing the data so that they can be processed by the next, or to produce files that will be required during the data reduction process.

The data reduction cascade can be summarized as follows:

- Data Processing
- Calibration frames production
- Data Reduction Preparation
- Data Reduction

The Data Processing is the Pipeline step where the data are corrected for gain, oriented all in the same way (blue to red going from left to right and bottom to top), trimmed and corrected from the overscan.

During the Calibration Frame Step, master bias, master dark (if needed) are produced from a set of biases and darks. The master flat is also created.

The Data Reduction Preparation Step is the series of tasks where the orders are located on the CCD. The wavelength dispersion relation is also determined during this step.

The Data Reduction Step is the final step of the execution of the PIpeline. All the calibration frames are combined with the science frame, and a one dimensional, wavelength calibrated spectrum is produced.

3. Scientific requirements for the Pipeline

The next table summarises some of the goals that must be reached by the pipeline in order for us to know that the milestones have been reached. This is not an exhaustive description of all that the pipeline will need to achieve, but merely a short description of the crucial goals the pipeline is required to achieve.

Resolving power	LR	13480/14850
	MR	37890/42850
	HR	71960/63230
	HS	60610/64810
Accuracy	LR	
	MR	
	HR	
	HS	0.001 pixel
Order detection	All modes	Order separation greater than 11 arcseconds or 5 pixels.
Wavelength Calibration	All modes	Merged scientific spectrum corrected for blaze function

4. Status of the HRS Pipeline

The HRS pipeline is currently under heavy development. The aims are to have the 4 HRS modes, namely Low Resolution, Medium Resolution, High Resolution, and High Stability supported.

As of 15 October 2015, the prototype HRS Pipeline is able to reduce data under a limited number of use cases and dates. The pipeline has not been thoroughly tested, does not support high stability mode, and is not guaranteed to produce usable data.

5. Pipeline Details

This section describes in more detail what the pipeline will actually be doing.

1. Processing the Data Frames

During that phase of the reduction cascade, the following steps are applied to the data:

- Correction of the overscan
- Trimming of the frames
- Correction of the amplifiers' gain
- reordering of the data, so that wavelength go from blue to red in the following way: bottom to top and left to right
- subtraction of the masterbias

2. Processing the Calibration Frames

During this step of the reduction cascade, the following Master Calibrations are produced:

- Master bias
- Master dark (if required)
- Master flatfield

3. Finding the orders

- Normalizing the master flatfield
- Creating the Order Frame, this step may have to be interactive in some cases

4. Wavelength Calibration

This step allows the transformation of the frame, from a (pixel, intensity) grid, to a (wavelength, intensity) one.

- Removing of the curvature
- Creation of a model of the instrument
- Extraction of the ThAr lines, with/from a linelist
- Creation of a wavelength solution for each order

- Iteration over the solution, to find the best one

5. *Reduction of the science frame*

This is the last step of the data reduction cascade.

- Use all the Calibration produced during the previous steps to produce a wavelength calibrated science grade spectrum, that has been corrected of most, if not all, instrument and telescope effects.
- Optimal removal of the sky features in the science frame.
- Merging of orders into a single spectrum
- Flux (spectral shape) calibration using standard star spectra

6. Man power estimation

This section estimates the workforce needed in order to meet the requirement of the HRS Pipeline User Requirements. It describes, section by section, following the plan of the HPS, who is needed to achieve the goal set. There are two kind of skills required : a Computer Scientist (CS) and an Astronomer (A). They may be more than one person, and in fact, it is very advisable that they are.
The order in which the required person is listed indicates who should take the lead on a given task.

1. *Work definition*

The Astronomer's work will be to define which algorithm should be used. He will be in charge of providing the most efficient ones and, if needed, a panel of algorithms to be used. He will be in charge of defining which mathematical tools are to be used, and to provide a way for the Computer Scientist to implement them. He will be in charge of ensuring the implementation have been properly done.

The Computer Scientist will be in charge of implementing the methods chosen by the Astronomer. He will do so following the Software Development best practices so that the pipeline can be efficiently maintained and developed. He will be tasked with defining the software requirements to run the pipeline, and will make sure these are translated in terms of hardware requirements.

2. *HRS Pipeline General Functionality*

1. Platform

- a) : CS for packaging, in order to have a pipeline that will be easily installable
- b) : CS + A to ensure modularity
- c) : CS to make sure the pipeline is available to all user who needs it.

2. Input Data

- CS + A : A shall provide specific details concerning the FITS format, specifically which keywords are to be used, CS will integrate these requirement into the pipeline.

3. Output Data

Same as for Point 2 above.

4. User interaction

- CS + A : CS to write an easy to use software, and A to provide informations on what As are usually expecting from a software.

5. Prototype Pipeline

- A : The prototype pipeline will be an early version of the HRS pipeline. It will work on a very limited set of data, with limited functionalities. Its final product will be reduced data, as defined in Sect 2 of the HRS Pipeline Spec

6. Command line and scripting

- CS + A : A knows how astronomers use a pipeline, and will transfer this knowledge to the CS. The CS will then implement the scripting and command line functionalities accordingly

7. GUI operation

- CS: This step has to be done using the appropriate Software development best practices

8. Automatic Mode

- CS + A : CS has the knowledge about how to make the pipeline run. He also has the knowledge to make sure the pipeline fails in a useful way, if/when it happens. A will let the CS know what steps need to be retried after a pipeline failure.

9. Individual Steps

- A + CS : each step will be defined by A, and CS will implement them so that each task is independent but easily serialisable.

10. Individual Steps mode user guidance

- a) A + CS : cf Sect 9
- b) A + CS : Complete documentation on what the software does, how it does it, what must be provided for each task, will be provided and kept up to date.

3. HRS Pipeline Technical Functionality

1. Classification and Calibrations

- A + CS : A will define what needs to happen with/to the data in terms of mathematical/physical processing. CS to implement.

2. Data Products

- A + CS : A to define what are the intermediate products that need to be kept. CS will make sure that the products are usable for the next steps. In particular, this will mean

making sure that the FITS header have been updated properly and contain the informations needed for the following steps.

3. Feedback

- A + CS : the pipeline must be verbose. A will define what information is needed so that the user knows things are going well (or not). CS will then implement it and will write the interfaces according the software development best practices.

4. Calibration Data Processing

A + CS : A to provide the mathematics needed to perform each task. And will also provide the most useful algorithms for each. CS will then implement them, and will make sure that flexibility is built in, so that adding more choices of data reduction recipes is possible,

5. Science Data Reduction Frame

- Cf Sub-Sect 4

4. Task breakdown

Following the block diagram in Section 3.3 of the HRS Pipeline Spec, the following man power split is defined:

1. User Interface : CS
2. Application Manager : CS
3. Data Cruncher : A
4. Data Storage : Information Technology Department
5. Data Display : CS

5. Summary of workforce needs

In many ways building a pipeline is never ending work. However, we can estimate the time requested to produce a working pipeline, that will deliver data that can be used directly to produce science, knowing how much work is needed, what has already been made, and what needs to be done.

Should we hire a competent computer scientist / programmer, he/she will need to work full time for 6 months, in order to complete the missing gaps of our current incomplete prototype pipeline developed by Steve Crawford and make it supportable and distributable. The most pressing issues are the development of robust and efficient routines for the wavelength calibrations, as well as robust order-merging procedures. Then, there is a dire need for proper and up to date documentation of the code, not to mention the packaging of the pipeline.

During these 6 months, however, we would also need an Astronomer available nearly full time for the process to progress efficiently.

These are not resources that we have available currently, unfortunately, particularly the Astronomer.

So a more realistic plan regarding funding and time would be to dedicate 12 months, starting as early as possible with the recruitment, to acquire a competent computer scientist / programmer, and have him/her work alongside an existing astronomer dedicating a smaller share of his/her time, to make sure we can produce a working pipeline. This mode is slightly slower, but avoids the need to hire a new (more expensive) astronomer. This plan is part of the proposed SALT Survival Mode budget for 2016/2017.

We also note that the skills acquired by the CS during the process should then be used to further develop pipelines for the other SALT observing modes. Developing RSS LS, MOS, FP, Polarimetric pipelines would take another 12 months. RSS/NIR needs will have to be looked at in the future.

Personnel	Time needed for HRS pipeline
Computer Scientist	12 months FTE
Astronomer	3 month FTE