MNRAS **000**, 1–22 (2016) Preprint 18 March 2016 Compiled using MNRAS LATEX style file v3.0

The GALAH survey: The data reduction pipeline

Janez Kos,[1,2]★ Jane Lin,[3] Tomaž Zwitter,[2] Maruška Žerjal,[2] Sanjib Sharma,[1] Joss Bland-Hawthorn,[1] Martin Asplund,[3] Tim Bedding,[1] Gayandhi M. De Silva,[4,1] Ken C. Freeman,[3] Sarah L. Martell,[5] Katharine J. Schlesinger,[3] Daniel Zucker,[6,7]

[1] *Sydney Institute for Astronomy, School of Physics, The University of Sydney, 44-70 Rosehill Street, NSW 2016, Australia*
[2] *Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia*
[3] *Research School of Astronomy & Astrophysics, Australian National University, ACT 2611, Australia*
[4] *Australian Astronomical Observatory, North Ryde, NSW 2133, Australia*
[5] *School of Physics, University of New South Wales, Sydney, NSW 2052, Australia*
[6] *Department of Physics and Astronomy, Macquarie Univeristy, Sydney, NSW 2109, Australia*
[7] *Research Centre in Astronomy, Astrophysics & Astrophotonics, Macquarie University, Sydney, NSW 2109, Australia*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present data reduction procedures used in the GALAH survey, utilizing the multi-fibre spectroscope HERMES and the 3.9 meter Anglo-Australian telescope. GALAH is a unique survey, targeting 1 million stars brighter than magnitude V=14 with a resolution of 28,000 and a goal to measure abundances of 30 elements. We outline the design and present all the steps in the Iraf-based reduction pipeline. The pipeline takes advantage of existing Iraf routines and other readily available software in order to be simple to maintain, tested and reliable. A radial velocity and stellar atmospheric parameter estimator software is also presented, as it is the last step in the reduction pipeline, before the data is released to collaborators. Together with the description of the pipeline we also provide quality analysis of the GALAH spectra; level of fibre cross talk and scattered light, resolution across the field, sky spectrum properties, wavelength solution reliability and the radial velocity accuracy.

Key words: spectroscopy: instrumentation – spectroscopy: methods – spectroscopy:surveys – spectroscopy:stars

1 INTRODUCTION

There is no rigorous recipe on how to reduce multifibre spectra. Multifibre spectrographs are sophisticated and complex machines and users are not expected to understand them in details, so they rely on the pipelines provided by the creators of the instrument. All such instruments have a dedicated pipeline available to the users (e.g. Blecha et al. 2000; Dominquez Palmero et al. 2014; Luo et al. 2015; Nidever et al. 2015), which is most often the pipeline that is used for large surveys as well. It is important that the reduction pipeline is well documented. The user must know which steps were made in the reduction and at what precision certain calibrations were made in order to understand the data the user is working with. Real value of any analysis highly depends on the reduced spectra quality and its properties.

The GALAH (GALactic Archeology with Hermes) survey (De Silva et al. 2015) utilizes HERMES, a multi-fibre spectrograph of the Australian Astronomical Observatory (AAO) 3.9 m Anglo-Australian Telescope (AAT) and the 2dF fibre positioner. The spectrograph produces high resolution spectra in four wavelength ranges or arms. This allows the GALAH project to measure the abundances of 30 elements in high quality spectra of 1 million stars. HERMES was built with the GALAH project in mind and GALAH is the first major survey being conducted with this spectrograph (Sheinis et al. 2014). The fibre positioner system called 2dF, however, is older and used by several other multi-fibre spectrographs. They all use a common reduction pipeline called 2dfdr (AAO software Team 2015). A reduction pipeline within the 2dfdr environment was written also for the HERMES, but it lacks many important features and proved to be often producing results of poor quality when used by the GALAH collaboration.

We therefore produced an alternative pipeline to be used by the GALAH survey. It is based on preexisting Iraf routines in order to be simple to maintain. Iraf has been widely used in the astronomy comunity for more than two decades. The routines we use have therefore been extensively tested and proved to be reliable (Tody 1986).

★ E-mail: jkos@physics.usyd.edu.au

Arm	λ_{min} / Å	λ_{max} / Å
Blue	4718	4903
Green	5649	5873
Red	6481	6739
IR	7590	7890

Table 1. Wavelength ranges covered by each arm of the HERMES spectrograph.

The Iraf pipeline is now the de-facto pipeline used in the GALAH survey. It is regularly updated and improved, but remains an inside project, not intended for a general HERMES user.

This paper briefly reviews the observational part of the GALAH survey in the section 2 and the motivation to write a new reduction pipeline in Section 3. The main part of the paper is the description of the pipeline, explained step by step in Section 4. After that, in Section 5, we describe the organization of reduced data. As a part of the reduction pipeline we also run another pipeline, a simple estimator of the radial velocity and basic parameters of stellar atmospheres. This is discussed in Section 6. Byproducts of the pipeline like linelists and resolution maps are collected in the Apendix.

2 OBSERVING PROCEDURE

Observing procedures in the GALAH survey are described elsewhere (Martell & the GALAH collaboration 2016; Sharma & the GALAH collaboration 2016), so we only give a quick review of the relevant parts here.

HERMES is a high resolution, multi spectra, spectrograph simoultaneously observing four wavelength ranges, called arms. The designed resolving power is 28,000 in a normal mode or 50,000 in a slit-mask mode. The normal mode is used for the GALAH survey. This gives a resolution element size of 3.5 pixels on each of the four CCDs. The observed wavelength ranges are given in Table 1. Due to the design of the spectrograph the wavelength ranges vary for few Ångstroms with the position of the spectrum on the CCD.

The GALAH fields for the whole survey are defined by a fixed pattern across the southern sky (Sharma & the GALAH collaboration 2016). The field configuration file contains all the relevant information about the stars that can be observed in each field, but not necessary at once. The observer selects a field from a list of observable fields at a given time and feeds it into the target allocation software. This software calculates what arrangement of fibres is most optimal to observe as many targets in the field as possible. It also allocates sky fibres and selects guiding stars. If the observer is satisfied with the configuration, one sends the configuration to the 2df fibre positioner robot that will actually move the fibres into the correct positions.

Fibres attach to a metal plate with magnets. There are two plates called plate 0 and plate 1 facing away from each other. The whole setup can be tumbled, so the fibres can be positioned on one plate while a separate set of fibres are receiving light from the telescope on the other plate.

There are 400 fibres on every plate that can be positioned by the 2dF robot. 392 of these are regular fibres and 8 are fiducial fibres used for guiding. Fibres on each plate on the telescope end are numbered from 1 to 400 with so called pivot numbers. Bundles of 10 fibres are then led into the spectrograph. Such group is called a slitlet. In each slitlet is an optical element that changes the focal ratio and as a result inverses the numbering order within each slitlet. The fibres on the spectrograph end are again numbered from 1 to 400 (now called fibre numbers), but because of the inversions, these numbers do not correspond to pivot numbers. During the reduction we ignore the fiducial fibres, because they do not contribute any signal to the image, so we only have to deal with 392 spectral traces. When apertures are fitted to the traces they are numbered, this time from 1 to 392, so the aperture numbers do not correspond to either fibre or pivot numbers. See Section A and Table A1 for a conversion key between the three numberings conventions.

At the beginning of each night observers take a series of bias frames for each ccd. These will be used for the reduction of the following images and do not have to be repeated later. In the mean time, the first two fields are given to the fibre configuration software and the fibres are positioned on the two plates. After the first field is observed, plates are switched and the second field is observed, while another field is configured on the now available plate.

Each field is usually exposed three times for 20 minutes per exposure. The observer can decide if more exposures are needed or if they should be aborted due to bad weather, for example. Before or after the science exposures are made, one exposure of the arc lamp and one exposure of the flat lamp is made with the same fibre configuration as for the science exposures.

Not all fields require 1 hour of exposure times. During the twilight we observe bright fields that require less exposure time. Special programs like observations of standard stars or K2 followup also use different exposure times.

The images created by four CCDs are kept in separate folders and are numbered by the run number, following the consecutive exposure done in that night. The fibre configuration is also saved, as it is the only record that shows which star was positioned in which fibre. It also stores the position of all fibres, including sky, fiducial, and unused fibres.

The rules, exceptions and patterns described above guided the design of the reduction pipeline.

3 MOTIVATION

It was originally intended that the GALAH survey data should be processed by the AAO's 2dfdr software. But it lacked specific functionality that needed to be optimized to cope with with problems unique to high resolution spectroscopy. Thus a more efficient and more specialized code was needed. A decision was made to write a second pipeline, produced and maintained by the GALAH collaboration for the purpose of the survey. Such a pipeline is under the control of the GALAH team, so any requirements can be implemented quickly.

The pipeline is only required to work with GALAH spectra that are all taken in the same manner, as the data have a predictable format. It is not meant to be released as

a pipeline for general use. To ensure the pipeline was fast to develop, easy to maintain or modify and reliable, we used preexisting and already tested Iraf routines as much as possible. For procedures not available in the Iraf packages, we, again, used existing solutions before writing our owns. Because of the tight dependence on Iraf, we call this pipeline simply the Iraf pipeline. The whole pipeline therefore contains only ~2000 lines of code. The cost of this approach is a slow execution of the code. The reduction (including radial velocity and stellar atmospheric parameters estimation) of the whole GALAH dataset (at the time of writing data for 200,000 unique stars or 600 fields) takes almost 2 weeks on a top end desktop PC. The main limitation here is Iraf's use of many read/write operations on files on the hard drive have to be done at each step. However, other advantages of Iraf were judged adequate for our needs.

The pipeline must be fully automatic, requiring no human interaction to reduce the data. This was achieved in the greatest extent possible where any further automation would require disproportional effort to implement.

It is also desirable that the same code can be used to reduce spectra from all four arms. This simplifies the pipeline significantly. This was achieved in the presented pipeline, except for a few steps that are skipped entirely for spectra from some arms, because they are not needed.

4 THE PIPELINE

This section describes the major steps made in the pipeline. They are presented in a logical order but they are not necessarily executed in the same order, as sometimes several iterations are made and some steps are performed at different instances.

The same pipeline is used for all images, regardless what observing program produced them. Before they are fed into the pipeline, images are separated into groups, 4 for each arm and 2 for each plate, making 8 groups altogether. Each group of images is then reduced on its own without any interaction between the processes working on each group. Only images from one night are reduced at once. Any number of instances of the pipeline can be run in parallel, making the reduction process faster and more efficient. Usually we run between 16 and 32 instances at once, so 2 to 4 nights can be reduced at once.

4.1 Non automated routines

There are several steps in the pipeline that cannot be completely automated when some new data is fed to the pipeline. They do not have to be performed every time the pipeline is run; some are done once for all or are reviewed from time to time. Such occurrences usually relate to changed instrument calibration or new observing strategy.

Trace finding algorithm relies heavily on a given key which tells which aperture is positioned at which pixel. Spectrum traces are found by making a cross section along a specified column and locating the peaks. The position and separation between traces are not uniform and can also change with time when cameras are refocused or CCDs moved. A key was produced that gives the location of the trace centre for each aperture in the reference column. When traces are found in a new image, a global shift is allowed against the key and a small shift of 3 pixel for each individual aperture is also allowed. The full width at half maximum of the trace is around 4 pixels in comparison and the aperture width we use is 6 pixels. The key has to be made by hand and can be used as long as the geometry of the traces does not change significantly. This has only happened once by the time of writing. When the traces are found in the reference column, they are progressed to other columns with the same criteria that the deviation from the key is 3 pixels at most. This guaranties that the apertures are positioned on the correct spectra every time and that the trace does not jump between the neighboring spectra.

A lot of time was invested into building the linelists themselves for the wavelength calibrations (see Section 4.7 for a detailed description). The linelists themself are not enough to make the wavelength calibration. The method also expects to be given an approximate relation to transform pixels into wavelengths before the arc lines are identified. A large number of spectra had the wavelength calibration done manually to establish this relation. Humans are much more susceptible to notice mismatched lines than computers and the established relation is used as a proxy for the wavelength solution. We monitor the wavelength solution from time to time to see if the proxy relation has to be updated for new data.

An important parameter that was found by eye is the threshold in the cosmic rays removal algorithm. If set too low, parts of the useful data will be tagged as cosmic rays which will have a huge impact on the quality of the spectra. If set too high, the cosmic rays will not be removed. Different types of images (normal exposures, fibre configurations with only one bright star, exposures with low signal due to cloud cover, etc.) were ran through the cosmic removal algorithm and inspected by eye. The thresholds were hard coded to give the optimal result in every type of image we imagined to check.

A large number of parameters, like orders of fitted polynomials, were determined by reducing a few images by hand and playing around with the parameters to get the best results. They are explained in their appropriate sections.

4.2 Cosmetics

With each set of science frames, one flat frame and one arc frame are taken, either before or after the science frames. Flat frame is a so called flap flat, produced by a system of lights and reflectors shining directly onto the fibre ends. An actual flat illumination is not guarantied, even more, the light enters the fibres at different angles and the amount of light that actually enters the fibre depends on the fibre position. A dome flat (a flat field produced by the light reflecting off a screen mounted on the dome) and a sky flat are taken from time to time, but not every night. These two give an illumination of fibres that is much closer to the ilumination the stars produce.

One flat field is created from all flap flats taken in one night for each arm and each plate. This averages out the uneven illumination, as the fibres are repositioned approximately every hour. This step produces evenly illuminated and consistent flat to use in the pipeline. See also Section 4.8 for the analysis of the sky subtraction process, which is

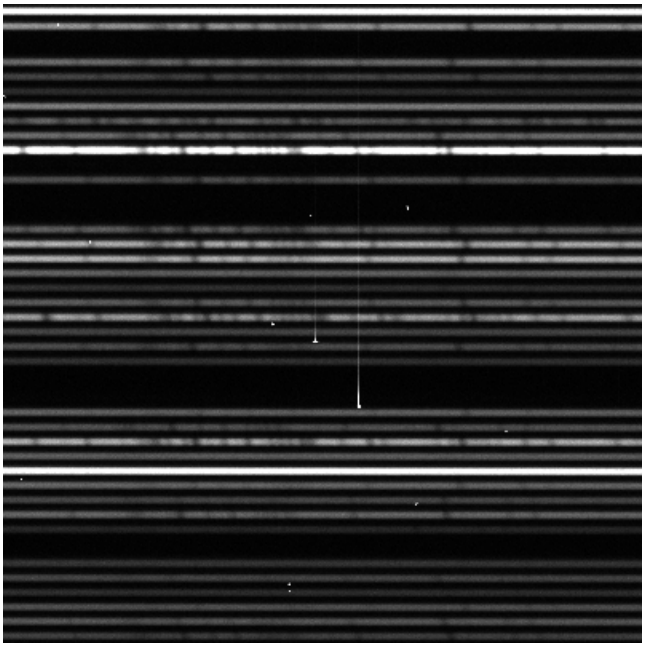

Figure 1. An example of two vertical strikes in a spectrum made on February 8th 2015, run 39, green arm. Horizontal lines are traces of stellar spectra with clearly visible stellar absorption lines. Two strikes are visible in the centre. A few cosmic rays are also visible. Size of this image excerpt is 400×400 pixels.

highly dependent on correct flat field reduction. Success in removing the sky emission lines is a prove that averaging several flap flats produces a usable flat field image.

A flap flat field does not illuminate the whole CCD uniformly, so we abstain from correcting the images for a response measured by a flat field in a conventional way. We also do not use dark frames. CCDs in the HERMES spectrograph are cooled to levels where the dark current is negligible. The first step in the reduction process is then bias correction. We use bias frames for that, where we verify that the bias levels are the same in the overscan of every image and in the actual bias images.

Our science exposures suffer from three cosmetic aberrations; cosmic rays, vertical strikes, and degraded columns.

Vertical strikes are suspected to be very energetic particles originating from the HERMES itself, but the source has not been confirmed yet. Although we suspect either optical glass or coatings. There are around 20 such events detected per 20 minute exposure in blue and green arm. There are none in the other two arms. A strike affects several pixels at the impact point that get saturated and the signal blooms into the pixels in the same column. The signal continuing from the blooming part gets weaker with distance from the impact point and degrades into undetectable levels after 500 pixels in most extreme cases and around 80 in a usual case. This means that a single event impacts many spectra. Some vertical strikes are persistent between images and cannot be erased by a reasonable number of pre-flushes of the ccd.

Fortunately a significant part of any strike is removed by the cosmic rays removal procedure. What is left is the centre, if it is too large, and the weak tails of the strikes that do not have enough contrast to be detected by the cosmic rays removal algorithm.

Degraded columns are only present in the IR arm on the first ADC (analog to digital converter). There are 6 such columns (columns numbers 3262, 3270, 3271, 3308, 3324, and 3347) and the signal in them is replaced by linearly interpolated flux from the neighboring columns. These 5 columns appeared suddenly on August 30th 2015 and have not changed the behavior since. They are all in the region of the spectrum uncorrupted by the telluric A band.

We are using a well known and tested method for detecting and removing cosmic rays (van Dokkum 2001) called LaCosmic. A python version of the code is used with minor modifications done by us. The method described in van Dokkum (2001) was designed to find and remove cosmic rays in images with no apparent patterns or symmetries, like plain images of astronomical objects. Images of spectra have a strong pattern in them, in our case horizontal traces of spectra. When a cosmic ray is detected, the flux in affected pixels can no longer be interpolated from the neighboring pixels without considering the pattern. In the release of the code used here, the affected pixels have flux interpolated from a box of 5×5 pixels in size. A median signal in this box is used to calculate the interpolated value in the pixel affected by a cosmic ray. We changed the box size to 7×3 pixels, with the longer side along the wavelength axis. This takes the obvious pattern into the account, as the interpolated value will be calculated from the pixels lying on the same spectrum trace. Keep in mind that the resolution element is around 3 pixels wide. The interpolated value is therefore calculated from the nearest two regions in the spectrum that can still be resolved, so the spectrum degradation in the pixel affected by a cosmic ray is minimal. This approach is still not perfect, as the spectra also have a predictable pattern in the axis perpendicular to the spectra traces; a distinct bell shape. Just changing the box shape, however, introduced a significant improvement.

In the blue and green arm, where we have vertical strikes, 0.073% of pixels are detected to be affected by cosmic rays by LaCosmic. This drops to 0.062% in red and IR arms. Both numbers hold for 20 minutes exposures we make for most science images. We are satisfied with the LaCosmic performance, so this is the only measure made to remove the comic rays. Eventhough we make 3 consecutive exposures of each field, we do not median them at any stage. Median of, say normalized spectra, is considered a good step to remove comic rays, but in our case it would fail to remove many vertical strikes, because they persist through several exposures. We therefore find this step redundant, but still give the user an option to combine the individual spectra in any way possible.

4.3 Tracing the spectra

Tracing of the spectra is done on an averaged flat field. This means that a single tracing is used for the whole group of images.

392 apertures must be positioned on the spectra traces in each image. A key is used to give the pipeline an approximate position of each aperture in one reference column. After the centre of the aperture is calculated at the reference column it is traced along the spectrum trace. The shape of

the trace is represented by a 7th order Chebyshev polynomial. The width of the aperture is fixed at 6 pixels and is centred at the peak of the trace.

4.4 Correction of optical aberrations

HERMES is designed to produce almost linear, equally spaced, 2 dimensional spectra traces. This is achieved almost perfectly. Optical aberrations are most prominent in the corners of the CCD, where the point spread function (PSF) takes an obvious elliptical shape. Because the PSF of the HERMES spectrograph is not predictable enough to account for any time changes and changes due to focus, temperature, etc., we choose to correct this aberration in an empirical way, by measuring the tilt of the ellipses and straightening the PSF, so it is narrower in the wavelength direction. This way we can improve the resolution in the corners. The PSF is adequately sampled by arc lines. We do this process for every aperture individually.

Each aperture is divided into 7 parallel subapertures. A spectrum is extracted from every subaperture. Arc lines are identified in every subaperture, but no wavelength callibration is done. Only the positions of centres of the lines in the pixel space are saved. It is then assumed that the centre of an arc line should have the same coordinate in every subaperture. This gives us a condition for a transformation that we are looking for. For every arc line, we want to transform the coordinates of the line centres in each subaperture to match the average coordinate:

$$\begin{aligned}(x_1,1) &\rightarrow (x_{avg},1)\\ (x_2,2) &\rightarrow (x_{avg},2)\\ &\vdots\\ (x_7,7) &\rightarrow (x_{avg},7)\end{aligned} \qquad (1)$$

where x is the coordinate in pixels along the wavelength axis. The second coordinate is just the subaperture number (see Figure 2 for the illustration).

With ~ 40 lines per spectrum and 7 data points per line we have hundreds of points to which we can fit the following transformation. The transformation is a linear skewness that can change along the wavelength axis and the change is described by a 3rd degree polynomial. Five iterations of the fitting process are made with outliers above 2.5 sigma rejected in each iteration. Such a transformation was found to be the simplest one that produced a significant and satisfactory correction.

A transformation has to be calculated for every aperture. The optics is stable enough that the transformation can only be calculated once per night and applied to all images. A single transformation could be used for many nights in a row, but due to the pipeline design where every night is reduced independently, we calculate the transformation once per night (an separatelly for each group of images). Arc and science images are transformed following the same transformation fitted in this step.

A visual representation of the resulting transformation can be seen on a transformed arc image in Figure 2.

4.5 Scattered light

Scattered light is an anomalous excess signal originating from the scatter of light in the optical components, detector itself and in a very small part the air in the spectrograph. While HERMES has exceptionally low scattering for a fibre spectrograph, it is still detectable and must be accounted for.

Organization of fibres into slitlets and a significant gap between each slitlet allows us to measure the amount of scattered light in individual images (see Figure 3). Gaps between slitlets are ~10 pixels wide. There are also 8 gaps that are almost twice the width between the slitlets, because a fibre that should be next to a gap is a fiducial fibre (used for guiding) and is therefore not connected into the spectrograph. No signal should be present in such gaps, so all the signal we detect there can be attributed to different abberations, mostly scattered light and fibre cross talk. Gaps due to dead fibres are not treated as such, because there is no guaranty that the fibre has exactly zero throughput. Such gaps are also narrower than the rest.

Apertures are always 6 pixels wide and centred on spectrum traces. We define gaps as regions at least 0.7 pixel away from the edge of any aperture. We use the Iraf's `apscatter` routine to measure, fit, and smooth the scattered light.

First the cross-sections of gaps perpendicular to the dispersion axis are made and the signal in 70 consecutive columns is averaged. This avergaing produces sufficient signal-to-noise ratio (SNR) to select the parts of the gaps where signal is lowest. This is achieved by a highly asymmetric sigma clipping algorithm (low rejection limit is 4 sigmas and high rejection limit is 1.1 sigma). These regions are assumed to sample the scattered light. The rest of the signal in the gaps will be attributed to fibre cross-talk in a later step. The scattered light is then measured and fit along the dispersion axis, but only in the regions defined in the previous step. A two dimensional 6th degree Chebyshev polynomial is used to represent the scattered light in both directions and is promptly subtracted from the image.

Figure 4 shows one of the crossections with a fitted polynomial and rejected points. 59 such crossections are made over the whole image.

4.6 Fibre cross talk

Fibre cross talk is the percentage of the light originating in one fibre and being detected in the region of the CCD (in the aperture) where the neighbouring fibre contributes its light.

Fibre cross talk is measured by utilizing the gaps again. Because fibre cross talk varies from fibre to fibre, it is impossible to measure it for each fibre individually, but we believe that the variations are small and predictable, so the fibre cross talk can be measured for each slitlet and interpolated to individual fibres. Amount of the fibre cross talk is independent of the spectrum in each fibre and can be measured in a flat field image. Two separate fibre cross talks must be measured; one for the leak from the preceding and one for the leak from succeeding fibres. In case a fibre is next to a gap we assumed no cross talk from that side.

For every slitlet we added two so called phantom apertures, one before the slitlet and one after the slitlet (as il-

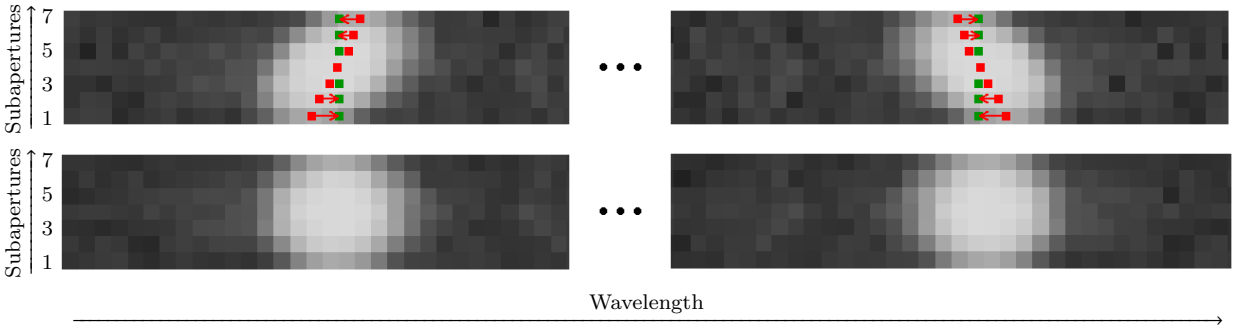

Figure 2. Example of two arc lines in the corners of a red arm image before (top) and after the geometric correction (bottom). Left image shows the shape of an arc line in the top left corner and right image shows it in the top right corner of an image. Notice that optical aberrations distort two arc lines into a differently oriented shape. Aperture is divided into 7 subapertures, seen here as lines of pixels. Calculated centre of the line in each subaperture is marked with a red dot and the centre of the line as measured over the whole aperture is marked in green. A transformation has to be found that moves the red points onto the green ones as illustrated by red arrows. The resolution in the bottom pannel is improved by 15% against the resolution in the top pannel.

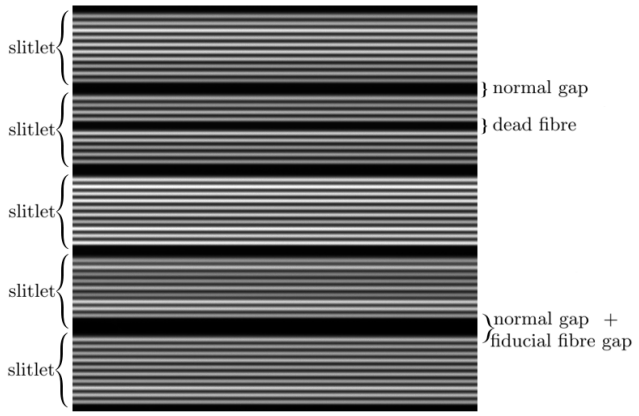

Figure 3. Slitlets and gaps between slitlets presented in a flat field image. Size of this image excerpt is 500×500 pixels.

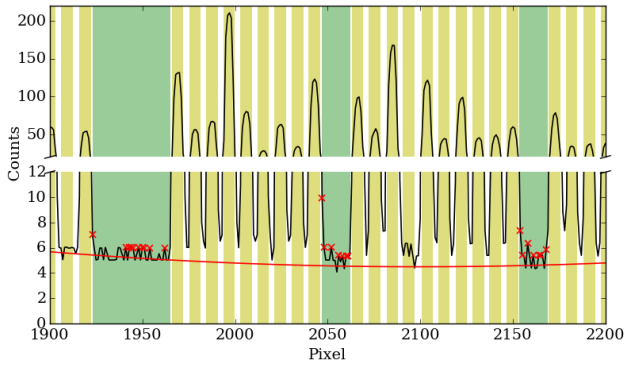

Figure 4. Fitted scattered light. Black line is the median of the spectrum in 70 columns. Yellow regions are apertures, and green regions are gaps. The fitted level of scattered light is the red line. Crosses show points in the gaps that were rejected by the sigma clipping algorithm. Only a small portion of the whole image is shown here.

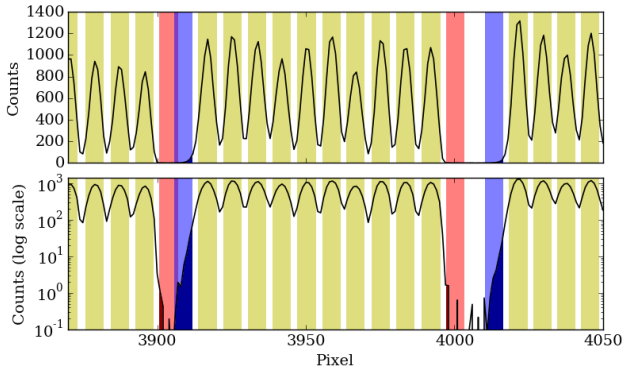

Figure 5. Cross-section of a flat field shows positions of the phantom apertures in the process of the cross talk measurement. Yellow are regular apertures, red are succeeding phantom apertures and blue are preceding phantom apertures, both placed in a gap. Part of the flux that counts as a cross talk is shaded in darker colours. The figure is presented in linear and logarithmic scales to give the right impression on the fibre cross talk values. Pictured is a corner of the flat field, where the fibre cross-talk is extreme.

lustrated in Figure 5), and extracted the spectra in all the apertures, including the phantom ones. Because the flux in phantom apertures should be zero, the actual measured flux is just the fibre cross talk from the neighbouring aperture. Dividing this by the flux in the proper aperture, we get the amount of fibre cross talk for this slitlet. Because each slitlet has a preceding and a succeeding phantom aperture, we measured both cross talk values. There are 40 gaps in every image. Some are rejected, because they include ghosts, as illustrated in Figures 6 and 7. The amount of the cross talk is interpolated with a 2 dimensional 6th order polynomial. This value is then used to subtract the fibre cross talk from every spectrum. The actual spectrum from the neighboring fibre that is subtracted is beforehand smoothed with a median filter with a boxy shape and 5 pixels wide. This prevents any remaining cosmic rays or vertical strikes to propagate into the neighboring spectra.

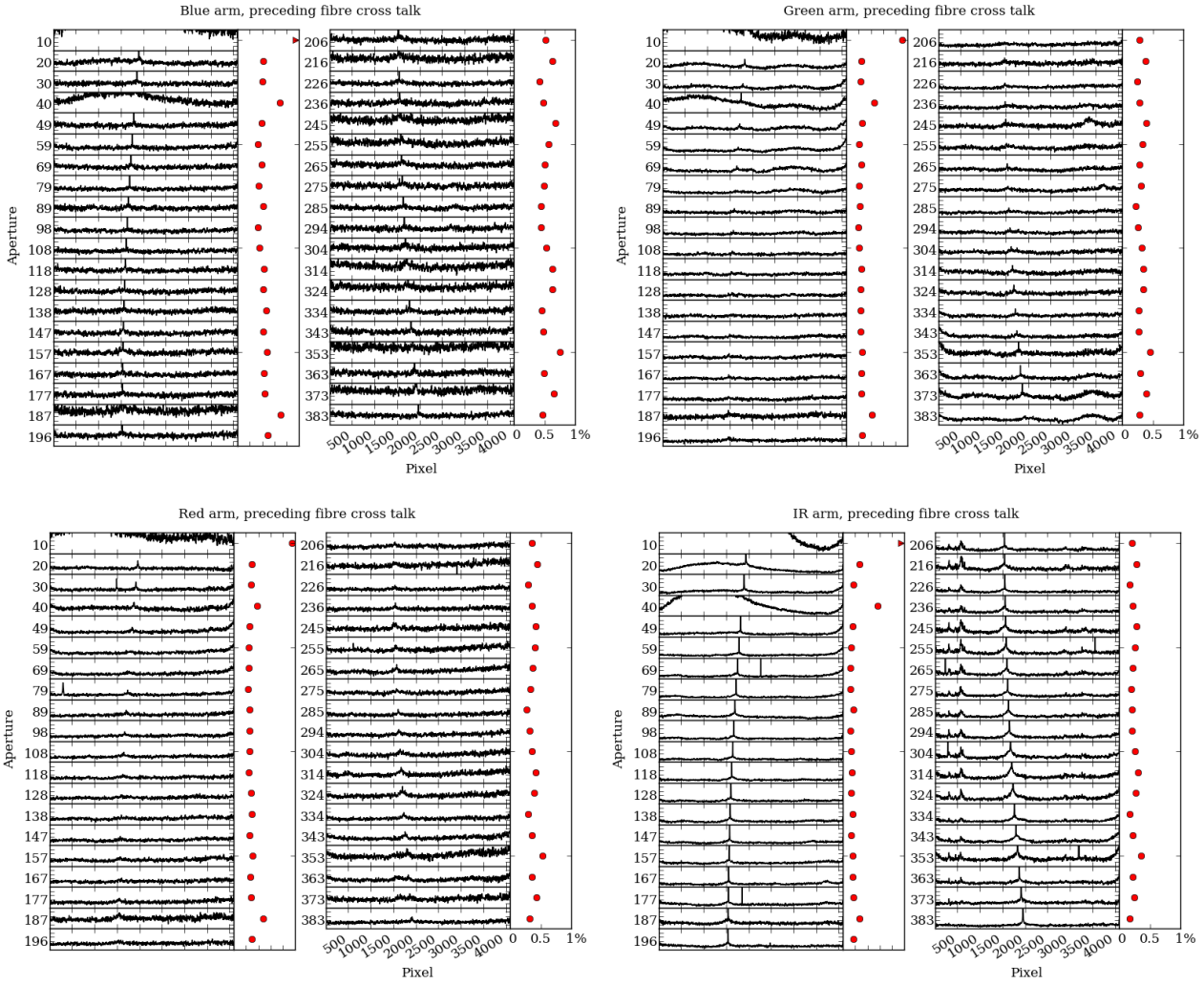

Figure 6. Variation of the preceding fibre cross talk as it depends on the wavelength (here in pixel scale) and fibre (here given as the aperture number). There are 39 gaps in which we can measure the fibre cross talk, hence 39 panels. Aperture number is the aperture next to the gap for which the fibre cross talk was actually measured. Every panel is 1% high. The adjacent panel with red dots shows the average fibre cross talk for each panel on the left from the dot. A spike in the middle of each panel is a ghost from the 0 order spectrum. It is removed from science frames, but not the flat field used here.

A formula to correct for the fibre cross talk is then:

$$\begin{aligned} S_F^{corr}(x) = S_F(x) \quad & - \Big[S_{F-1}(x) M_p(x,F)\Big]\, \delta(F,F-1) \\ & - \big[S_{F+1}(x) M_s(x,F)\big]\, \delta(F,F+1) \end{aligned} \tag{2}$$

$$\delta(F_1,F_2) = \begin{cases} 0, & \text{there is a gap between } F_1 \text{ and } F_2 \\ 1, & \text{otherwise} \end{cases}$$

where S_F is a spectrum in fibre number F, x is a coordinate running along pixels in the wavelength direction, and M_p and M_s are preceding and succeeding fibre cross talk maps like in Figure 8.

The amount of fibre cross talk highly depends on the aperture width which is an arbitrary parameter. We choose to have the aperture width fixed at 6 pixels. This means that we loose some of the signal that does not fall into the aperture, but at the same time keep the fibre cross talk low. A width of one pixel more, for example, would mean that the fibre cross talk is in the same order of magnitude as the signal gained by a wider aperture. The aperture width was chosen to keep the fibre cross talk lower than that.

Measured fibre cross talks prove that the cross talk changes smoothly from gap to gap, so we can safely assume that it is a result of spectrograph opticas and should change smoothly from fibre to fibre as well. This means that measuring the fibre cross talk in the gaps only is completely sufficient.

4.7 Wavelength calibration

Wavelength scale is calibrated through arc spectra of a Thorium-Xenon (ThXe) lamp. One arc spectrum for each arm is taken before or after the science exposures. It serves as a calibration for the science exposures observed with the same fibre configuration only.

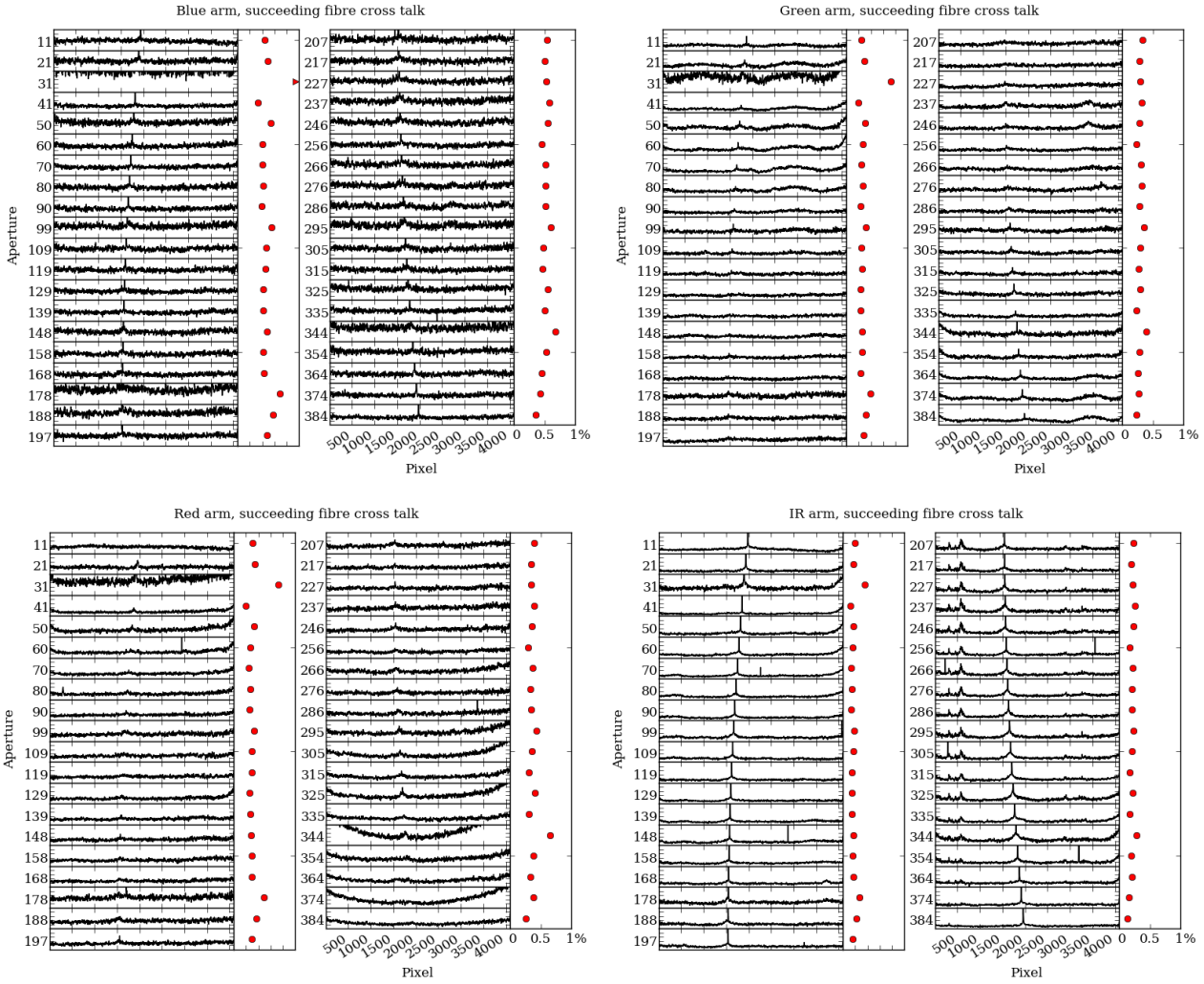

Figure 7. Variation of the succeeding fibre cross talk as it depends on the wavelength (here in pixel scale) and fibre (here given as the aperture number). There are 39 gaps in which we can measure the fibre cross talk, hence 39 panels. Aperture number is the aperture next to the gap for which the fibre cross talk was actually measured. Every panel is 1% high. The adjacent panel with red dots shows the average fibre cross talk for each panel on the left from the dot. A spike in the middle of each panel is a ghost from the 0 order spectrum. It is removed from science frames, but not the flat field used here.

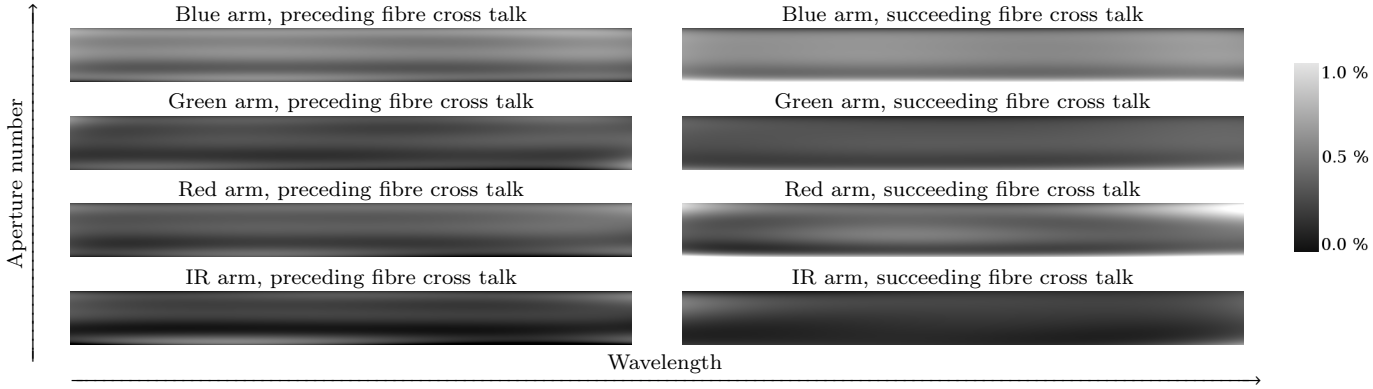

Figure 8. Fitted and smoothed fibre cross talks for every pixel and for every fibre. There are 4094 pixels per fibre and 392 apertures (8 fiducial fibres are not shown). The gray-scale shows the amount of cross talk and is normalized to 0 to 1% range.

4.7.1 The ThXe lamp and line identification

The dominant lines in our lamp belong to Xe, which is not widely used in the visual or near-infrared spectroscopy under 8000 Å. All the linelists we found in the literature lack the desired precision and are often inconsistent with each other. We also see lines in our arc spectra that do not have a matching line in the linelists. As is evident from the Figure B1 we have to use as many lines as possible, as there are not that many present. Thorium lines are in average much weaker than Xe lines, but they have well known wavelengths. However there are not enough Thorium lines to ensure a good wavelength calibration without using the Xe lines. The most notorious are the regions in the green, red and IR arms with no Thorium lines at all or only with one very weak Thorium line.

In order to assemble a reliable linelist, we made a set of special arc exposures with ThXe and Thorium Argon (ThAr) lamps. ThXe and ThAr lamps were turned on at the same time, so both sets of lines are exposed on the same image, removing the concerns about the stability and time variation during the exposure. 360 second exposures were made for blue, green and red arms. This is double the time we use for regular arc exposures. In the IR arm we only did a set of 120 second long exposures, otherwise some argon lines would get saturated and could potentially damage the CCD.

The ThAr+ThXe arcs were wavelength calibrated against the ThAr linelist provided in Iraf's `linelist/thar.dat` file (wavelengths for this linelist are taken from Palmer & R. Engleman (1993)). To assemble the ThXe linelist, the wavelengths of Thorium and Xe lines were measured in the calibrated spectra. We inspected each line by eye and excluded lines that have peculiar profiles like strong asymmetry due to blended lines and double peaks. We also omitted weak lines that are lost in the noise in regular arc exposures, but just visible in longer exposures of ThAr+ThXe.

The initial wavelength calibrations of the ThAr+ThXe lamp spectra done on Thorium and Argon lines have a root mean square of 0.0099 Å, 0.0054 Å, 0.0079 Å, and 0.0129 Å in the blue, green, red, and IR arm respectively.

This two step process is expected to introduce some systematic errors in the measured wavelengths. Figure 9 shows the discrepancies between our linelists and the literature. The large scatter is due to imprecisions in the linelists from the literature, where mostly Xe lines have wavelengths measured in low resolution spectra with a very limited accuracy. The shift is purely statistical. A similar amount of shift is seen when comparing lines from different literature sources, so having a similar shift against our linelist is expected. We further demonstrate that the shift is coincidental in Section 6 and Figure 14.

Linelists are collected in Appendix B.

4.7.2 Wavelength calibration

The apertures fitted to the flat field are used to extract science spectra and arc spectra with exactly the same parameters. There is some background noise in the spectra of arcs, which varies from fibre to fibre and from time to time. This noise can interfere with line identification and the measurements of the line positions, especially for weak lines which are crucial in our case.

To assure as consistent line identification as possible, every arc spectrum has lines identified with a range of different thresholds taken into the account. The threshold used to find the correct number of lines is then adopted. Detected lines are matched to the lines in the linelist using a proxy wavelength calibration done by hand on a small subsample of data and used for all the nights. The same function is then fitted to the detected lines. Lines that are more than 2.5 sigmas away from the fit get rejected. 5 such iterations are made and the wavelength calibration is in the end made on the best matching lines. We limit the number of rejected lines to 15% of the whole linelist. This limit is usually not reached.

The fitted wavelength calibration is monitored and if it deviates too much from the calibration function we found by hand, the spectrum is tagged so the analysis pipelines can skip it.

4.8 Sky subtraction

Out of 400 fibres, 25 are dedicated as sky fibres in every field during the field configuration step. These fibres can be different for each field. They are positioned as uniformly as possible across the field. Known stars are avoided, but not the galaxies or other objects, so it is possible that a fibre is positioned to an inconvenient location. Not all 25 fibres can therefore be trusted. It can also happen that less than 25 fibres are positioned, because it is impossible to configure the filed in a better way.

In the sky subtraction process first the sky spectrum is modeled across the whole field. It is than subtracted from all the spectra.

4.8.1 The sky spectrum

On a moonless night the most prominent features in the sky spectrum are air glow lines. They are very strong in the IR and red arms, there are only a few noticeable in the redward part of the green arm and none in the blue arm.

Whe nthe moon is up, the majority of the flux comes from the moonlight. Fields closer than 30° from the moon can and usually are avoided, so the moonlight contributes up to 120 counts per hour per pixel to the sky spectrum.

In addition, we observe an unexplained flux of in average 10 to 20 counts per hour per pixel (in blue and IR arm respectively). The spectrum of this contribution appears smooth. Despite the effort we were unable to identify the source. It does not show any telluric absorption lines, so it must originate from somewhere near the telescope.

In a few fields we observe emission lines coming from hot gas in the Milky way. In such a case the sky subtraction cannot be trusted, as the nebulae in the Milky way are much more structured than what can be sampled by 25 sky fibres. Fortunately the GALAH survey observes regions far away from the Galactic plane where such occurrences are very rare. The pipeline has no safety switch to tell the compromised fields from normal ones, but a spectrum before the sky subtraction is provided to the user, so the user can decide if the subtraction is still usable or not for his particular application.

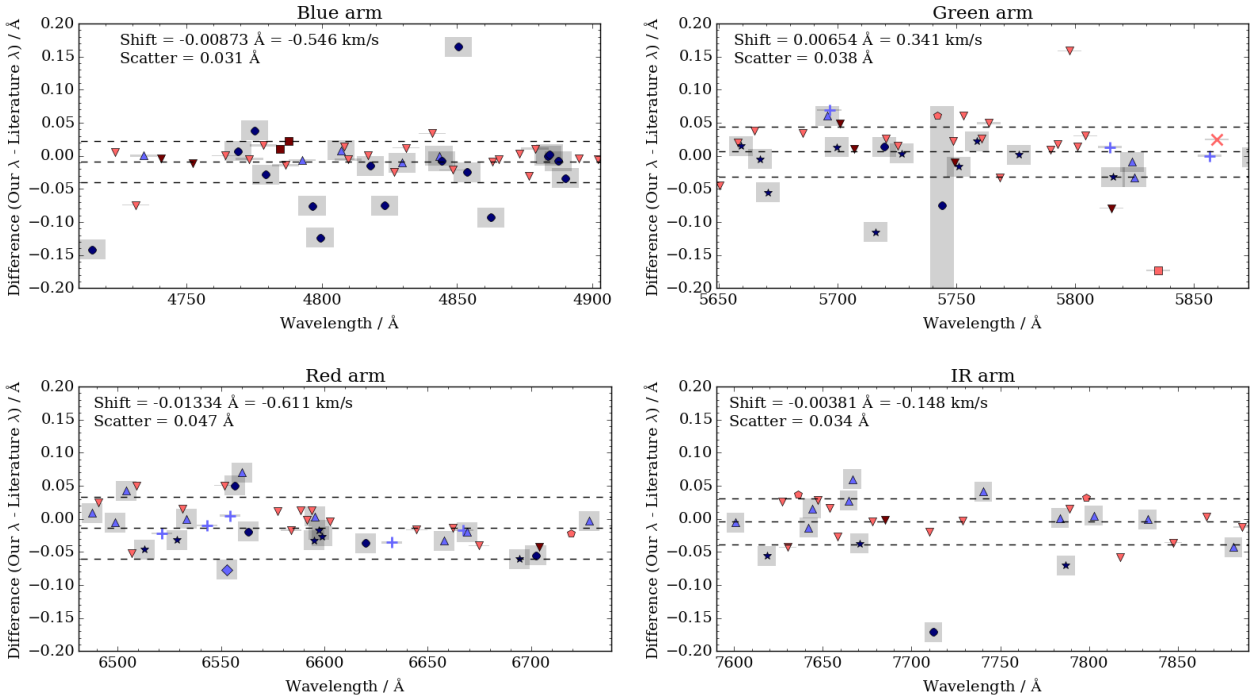

Figure 9. Difference between the wavelengths in our linelists and wavelengths of the same lines in the literature. Dark red symbols represent Th II lines, bright red Th I, dark blue Xe II and light blue Xe I lines. Shape of the symbol marks the literature source (●: Hansen & Persson (1987), ▼: Palmer & R. Engleman (1993), ▲: Humphreys & Meggers (1933), ■: Lovis & Pepe (2007), ⬠: Redman et al. (2014), ★: Humphreys (1939), +: Meggers & Humphreys (1934), ×: Zalubas (1976), ♦: Ahmed et al. (1998), ⬡: Zalubas & Corliss (1974) or respectively *a* to *k* in Table B1). Gray rectangles show the precision of the wavelength given in the literature. The largest rectangle in the green arm plot means that the wavelength is given with 1 decimal place, next size is 2 decimal places. Average shift and dispersion are plotted with dashed horizontal lines. Lines that we did not find in the literature and have only the nearest line given in the table B1 are excluded from this calculation.

4.8.2 Subtraction

With 25 sky fibres, including possible ill-positioned sky fibres we have to interpolate the sky spectrum over the whole field with the diameter of 2°.

In addition we must be aware of the changing resolution across the wavelengths and fibres. The later is solved in a sufficient way by correcting the optical aberrations. Only a small variation in the order of 3% remains (see Figures C1 and C2). This is less than the relative noise of the sky emission lines, so taking it into account is not necessary. The resolution is changing more along the wavelength axis, but this change is the same for sky fibres and object fibres so the subtracted sky spectra always have the right resolution at every wavelength.

In the first step we measure the fibre throughputs from a flat field. Excluding the dead fibres, the throughputs of fibres can be up to 50% different for different fibres (Simpson & the GALAH team 2016). To remind the reader, the flat lamps do not illuminate fibres evenly. The ilumination depends on the position and orientation of the fibre. On the other hand, a flat field is taken for every observed fibre configuration, so combining flat fields done for different configurations is enough to average out the variations in the fibre illumination. Fibre throughputs are therefore measured from am averaged flat field and are represented by 5th order Chebyshev polynomials, tracing the throughput at all wavelengths. It is assumed that the fibre with the highest throughput has the throughput equal to one and all the spectra are normalized according to this scale with their corresponding polynomials. The actual values of the throughputs are not relevant, because we do not intend to produce flux calibrated spectra.

All 25 sky spectra are then averaged with a min-max rejection of 3 sigmas. Spectra from ill-positioned fibres or any remaining cosmic rays, hot pixels, and vertical strikes are therefore removed from the averaged spectrum. This gives us a high SNR sky spectrum, so the subtraction will introduce less error.

In the next step the flux of the sky is mapped across the whole field. The median flux in each sky fibre is measured and the positions of the fibres on the plate are read from the fibre configuration file. The sky flux is mapped by fitting a plane to the measured median sky fluxes. The plane is represented by three parameters, tilt of the plane in two directions and the offset in the flux dimension. Fitting is done by minimizing the squared distances of points above or below the plane. Five iterations of the fitting are made where points more than 3 standard deviations away from the plane are rejected in each step. This removes any outliers like ill-positioned fibres. An example of such plane is shown in Figure 11.

We tried fitting more advanced functions, the next one being a warped plane, where the tilt of the plane changes along one direction. This proved to be too complicated to

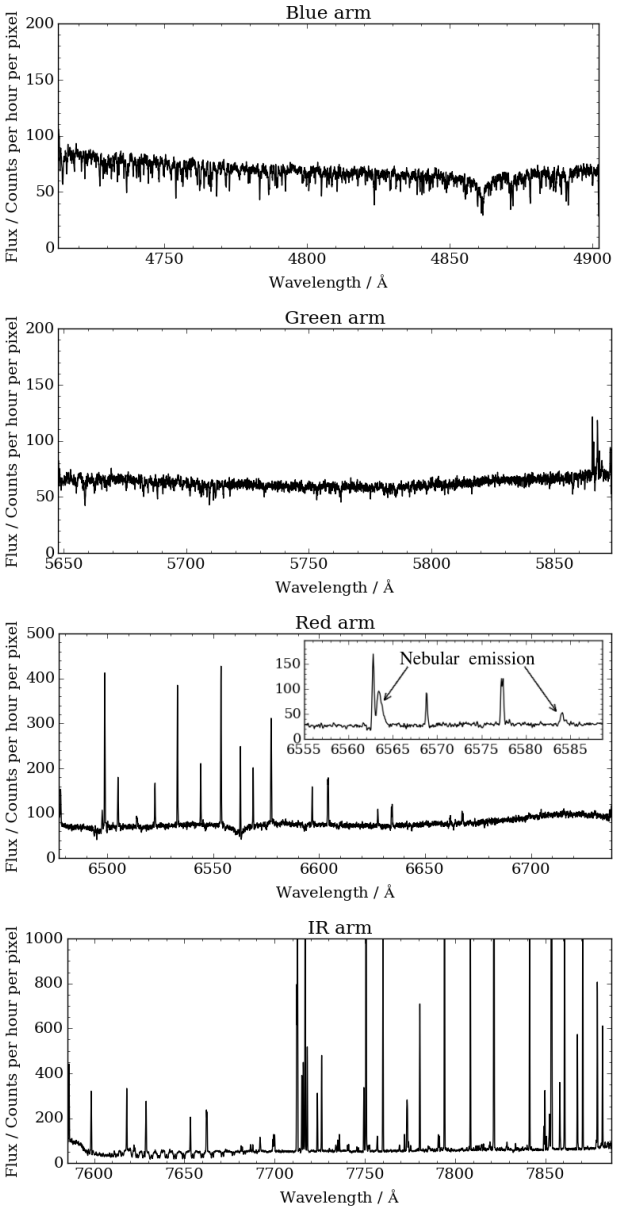

Figure 10. Spectra of a moon-lit sky on June 6th 2015. A typical spectrum of the night sky is shown for all 4 arms. The main contribution to the sky spectrum is in this case the moon-light. Almost all of the emission lines come from the air glow. The zoomed in region is from a different night than the other spectra, when there was no moon above the horizon and from a field positioned near the Galactic plane where nebular emission lines are present. Two marked lines are Hα and [N II]. All spectra are an average of all the sky spectra in one field. There is residual flux of around 10-20 counts in all the spectra. Its spectrum appears smooth and we could not find a good explanation for it.

fit automatically to 25 points if rejection of some points is allowed.

In the final step we take the position of each fibre, calculate the flux from the value of the plane at that position, scale the high SNR sky spectrum accordingly and subtract it from the spectrum for that fibre.

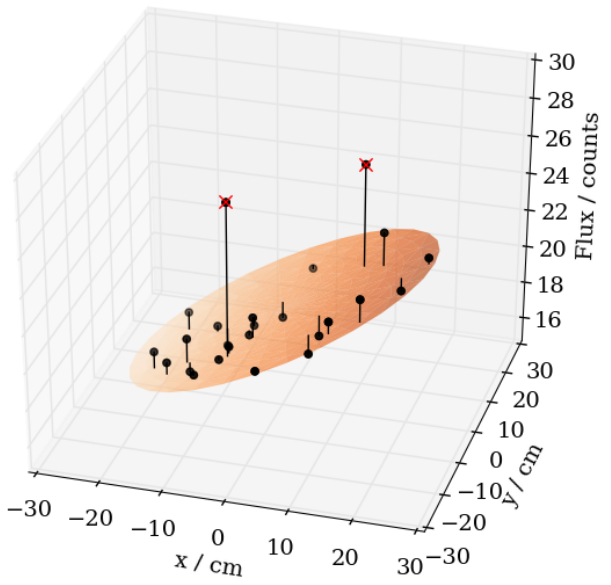

Figure 11. The process of fitting a plane to sky fiber fluxes. Measured fluxes are represented by black points. After the fitting two points were rejected and are crossed out in the image. The plane shows the size of the plate where fibres can be positioned. It is colour coded by the mapped flux ranging from 16.5 counts to 23.8 counts.

The pipeline cannot detect whether the sky subtraction algorithm is compromised by nebular emissions or bad weather, so the above procedure is done in the same way every time.

4.9 Telluric absorption lines removal

Telluric absorption lines removal usually requires additional observations of standard stars, which is time consuming and unfeasible for the GALAH. It also requires to model the absorption lines profiles across the sky and in time to properly account for the measured telluric absorption in science spectra. Because the GALAH bands are affected by the telluric absorption lines and have to be corrected regardless, we correct our data by fitting the telluric lines directly to observed spectra.

Telluric absorption lines are present in all four arms. However, only in red and IR arm are they strong enough to be detectable and corrected for. Part of the IR arm is in large part truncated by the A band, a series of O_2 vibrational lines. These lines are saturated and blended and are hard to correct for in a time efficient and automatic way. The IR arm wavelength range was selected in such a way that the oxygen triplet lines lie in the centre of the range. This means a large part of the A band must be included as well. There are no features essential to the GALAH in the A band. The only exception is the K I doublet, which lies in the less affected region with optically thin telluric lines.

Telluric absorptions are fitted with ESO's Molecfit software (Kausch et al. 2015; Smette, A. et al. 2015), release 1.1.1. Molecfit fits synthetic telluric absorption lines to spec-

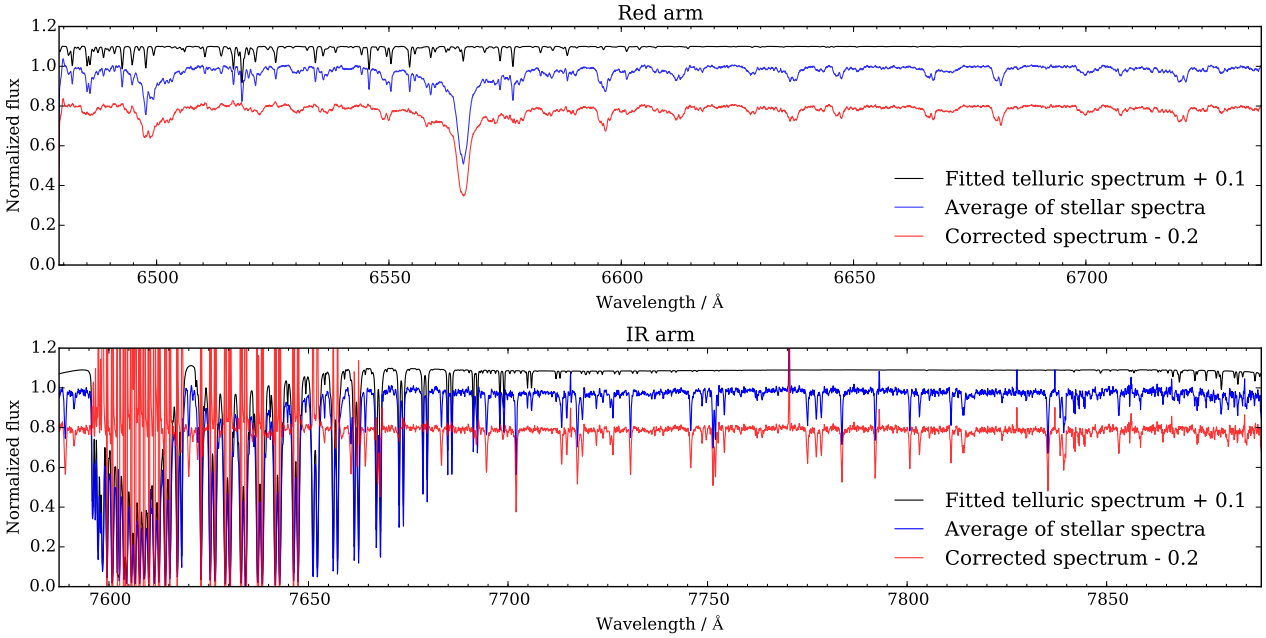

Figure 12. Telluric absorption lines are fitted to spectra. Plots show the fitted and corrected spectra in the red arm (top panel) and in the IR arm (bottom panel). The average spectrum of all objects in one image is plotted in blue. Black spectrum is the fitted telluric absorption spectrum. The corrected spectra are in red. The correction fails for lines in the A band that are optically thick but still works for the optically thin part of the A band. The plotted spectra are normalized for the purpose of this illustration. The process in the pipeline is done on non-normalized spectra.

tra. To fit the telluric absorption spectrum to one spectrum takes around a minute on a top-end desktop computer. It is therefore unfeasible to run Molecfit on every single spectrum. Instead we first average all the spectra of objects in one image and fit the telluric absorption spectrum only once. Our fields are observed at airmasses low enough that no telluric line will become optically thick in any part of the field. The exception is the A band, of course, which we do not intend to correct anyway. For each fibre we calculate the difference from the average airmass and scale the fitted telluric spectrum accordingly. The individual spectra are then divided by scaled telluric spectra. An example of fitted and removed telluric lines is in Figure 12

4.10 Barycentric velocity correction

The wavelength calibration is modified with the barycentric velocity correction. The correction takes the effect of the moon motion and rotation of the Earth into the account and calculates the projected velocity in the line-of-sight against the barycentre of the Solar system (Czesla 2015; Landsman 2015). Already available routines proved to be fast and accurate in performing this step.

4.11 Error spectra

Almost every step described above can be completed only with a final degree of certainty, even more so, because most of the parameters like scattered light or fibre throughput are measured from the spectra themself. Almost every step therefore introduce some uncertainty into the final spectrum. With the reduced spectra we also calculate the error spectra, giving standard deviation of each pixel in the spectrum. Because of the many steps performed in the reduction process this is no longer a poissonian standard deviation, but consists of a combination of components, of which some must be propagated through the pipeline when they are affected in consecutive steps.

5 ORGANIZATION OF REDUCED DATA

5.1 Individual spectra

Every reduced spectrum is named with a 16 digit number with an actual meaning:

$$\underbrace{140231}_{a}\,\underbrace{0043}_{b}\,\underbrace{00}_{c}\,\underbrace{125}_{d}\,\underbrace{2}_{e}\,.fits$$

a) Observation date in the format 2 digits for the year, 2 for the month and 2 for the day. This is the date when the night starts. It does not change during the night.
b) Run number.
c) Combining algorithm. 00 in this case means that this is an uncombined individual spectrum (one out of 3 exposures).
d) Pivot number.
e) Arm number. 1 is blue, 2 is green, 3 is red, and 4 is the IR arm.

This name is used instead of the star name or input catalogue number because it is unique. Not all stars that we observe have the input catalogue number. Such are callibration stars, some stars from the pilot programs, and programs that do not use the GALAH selection function, but share the same observing routine.

There is one fits file created per reduced spectrum. It contains four extensions. First one is the reduced spectrum itself, without any normalization. Second extension is its error spectrum. It gives the relative error. Third extension is the spectrum without the subtracted sky and with no telluric absorptions fitted. Apart from these two steps it is passed through the exact same steps in the reduction as the spectrum in the first extension. The fourth extension is its relative error spectrum.

The wavelength calibration is the same for all four extensions. It is linearized and in a standard fits format given by the NAXIS1, CRPIX1, CRVAL1, CRDELT1, and CUNIT1 for the number of pixels, the reference pixel, wavelength in the centre of the reference pixel, dispersion per pixel, and units respectively.

5.1.1 Combined spectra

Most often three consecutive 20 minute exposures are made for each field. This is not a rule. There can be more, if the observer concluded that the weather conditions were bad and more exposure time is needed, for example, or less. A flat field and arc frame are also done with the same fibre setup as the science exposures for each field. Such continuous block of exposures is assigned a number called the cob_id (continuous observational block identifier). A combination of the science exposures from this block make one observation that is used in the further scientific analysis. Individual uncombined spectra are usually not used. cob_id is simply the combination of 6-digit date marker and run number of the first exposure made in this block, regardless its type (flat, arc or science exposure). If the same field is observed in two parts with other fields being observed in between, two cob_ids will be assigned and it will be threated as two separate observations. The cob_id also changes if the same objects are observed with different plates. Even if the objects are the same, the field configuration is different, so such observations must be treated separately.

A new name is assigned to combined spectra:

$$\underbrace{1402310042}_{a}\underbrace{\_01}_{c}\underbrace{\_125}_{d}\underbrace{\_2}_{e}.fits$$

Instead of the date and run marker we used the cob_id here (marked with a). Note the difference in the c section of the name. Number 01 indicates that this is a combined spectrum and the method used for combining all the science exposures was the method number 1. This also implies that the cob_id is used in the first part of the name.

The structure of the fits file is the same as for the individual spectra with the addition of a list of spectra that were combined in each header.

Currently there are two combining methods we are using:

1: All the science exposures with the same cob_id are summed together. Error spectra are combined appropriately, so they still give relative errors.

2: All the science exposures with the same cob_id and all the exposures of the same object made at earlier times are summed together. Error spectra are combined appropriately, so they still give relative errors. The first part of the name remains the same; it is the cob_id of the last exposure combined. Even if there are exposures of the same object made at later times in the database, they are not added into this file.

6 RADIAL VELOCITY AND STELLAR ATMOSPHERE PARAMETERS ESTIMATION

After the reduction of spectra we run a code that gives the first estimates of the radial velocity and three basic parameters of the stellar atmosphere for each spectrum. The spectra are delivered to the collaboration only after this step is performed, so we present this pipeline together with the reduction pipeline. Radial velocities and stellar parameters are calculated using the blue, green and red arms only. Arm 4 is omitted as it contains significant amount of telluric contamination and fewer number of spectral lines.

Radial velocities (RVs) are calculated by cross correlating the observed spectrum with a set of 15 AMBRE model spectra (de Laverny et al. 2012). Each spectrum from each arm is individually cross correlated with the grid, one model at a time. The model set spans 4000–7500 K in temperature, with 250 K intervals. It has fixed $\log g$ and metallicity at 4.5 dex and 0 dex respectively.

The observed spectrum is first normalized by fitting the whole spectrum by univariate spline functions. Areas of prominent spectral features (such as the Hα and Hβ lines) are omitted from the fitting. Both the model and the observed spectra are then moved to zero level by subtracting their means and are multiplied by a window function to smooth out their edges. The normalized spectra are then cross correlated with the model grid to determine the RV from this particular model. The cross correlation peak is fitted with a 2nd degree polynomial and we adopt the peak of the fit as the cross correlation coefficient.

In calculating the final RV for the spectrum, only models which have cross correlation coefficient greater than 0.3 are included. The final RV for the arm is a combination of the good models weighted by their cross correlation coefficient. The final RV and its uncertainty for a star is the average and the standard deviation of RVs from three arms. If one arm has a RV lying more than two times further than the difference between the other two, it is excluded from the final RV. Figure 13 shows the compatibility between RVs from this work and RVs from literature for four well studied clusters. With the exception of M67, most stars are within literature errors.

Figure 14 shows the statistics for the differences between the RVs measured in each arm. There is a systematic shift between each arm. It does not correspond to the shift we see in Figure 9, concluding that the shifts given in Figure 9 are a consequence of imprecise literature values for a large number of lines. The nonzero RV differences in Figure 14 we attribute to systematic errors in the RV calculation scheme and the initial linelist we composed, which issues are propagated to every RV calculated.

Stellar parameters (T_{eff},$\log g$ and [Fe/H]) are derived by global fitting the observed spectra with a grid of 16783 AMBRE spectra (de Laverny et al. 2012). Tables 6 and 6 contain the wavelength and parameter ranges of this grid. The aim of this procedure is to provide accurate enough initial stellar parameters for the ionisation-excitation balance

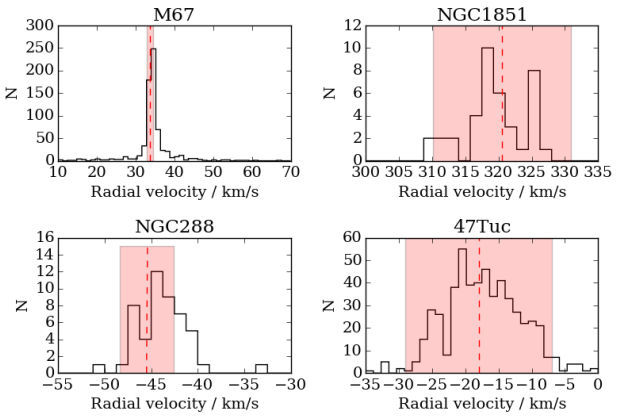

Figure 13. RV distributions of four well studied clusters. The red dashed lines are literature values, the red boxes are literature uncertainty ranges. Top left panel: M67, 716 stars. Literature RV: 33.64±0.854 km/s (Geller et al. 2015), RV from this work: 33.66±7.46 km/s (we adopt the standard deviation of our distribution as the RV uncertainty). Top right panel: NGC1851, 60 stars. Literature RV: 320.5±10.4 km/s, this work: 319.45±4.72 km/s (Harris 1996). Bottom left panel: NGC288. Literature RV: −45.4±2.9 km/s, this work: −43.92±2.62 km/s (Harris 1996). Bottom right panel: 47 Tuc, 662 stars. Literature RV: −18±11 km/s, this work:−17.46±6.35 km/s (Harris 1996).

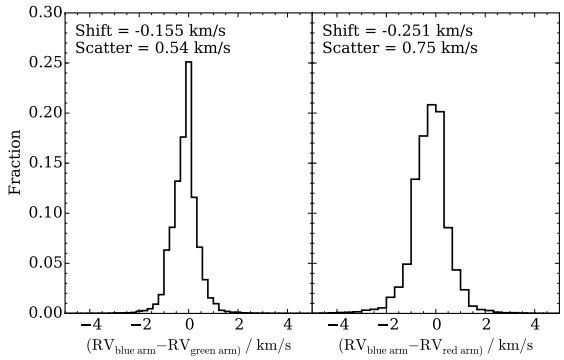

Figure 14. Difference between the blue arm and green arm radial velocity (left panel) and the difference between the blue and red arm radial velocity (right panel). Radial velocity is not measured from the IR arm. Shift from zero difference and scatter of the distribution are given in each panel.

pipeline in the analysis stage. This global fitting method is both robust and efficient, taking under a second for a star to converge.

The RV corrected spectrum is normalised by fitting 3rd (for blue and green arms) and 4th (for red and IR arm) degree polynomials to predetermined continuum regions. These regions are determined by examining high resolution spectra of the Sun, Arcturus and μ Leo. Normalised spectra are interpolated on to the common wavelength grid as the model. The ten closest model neighbours to the observed spectrum in Euclidean space are found. We then reconstruct the data spectrum using a linear combination of these ten

CCD	colour	wavelength / Å	sampling interval / Å
1	blue	4717 – 4887	0.0453
2	green	5651 – 5864	0.0546
3	red	6483 – 6729	0.0623

Table 2. Wavelength coverage and sampling for the model grid.

parameter	range	increments
$T_{\rm eff}$ / K	2500 – 8000	200 ($T_{\rm eff}$ < 4000) 250 ($T_{\rm eff}$ > 4000)
logg / log (cms^{-2})	−0.5 – 5.5	0.5
[Fe/H] / dex	−5 – 1	1 ([Fe/H] < −3) 0.5 (−3 < [Fe/H] < −1) 0.25 (−1 < [Fe/H])

Table 3. Parameter summary of the model grid.

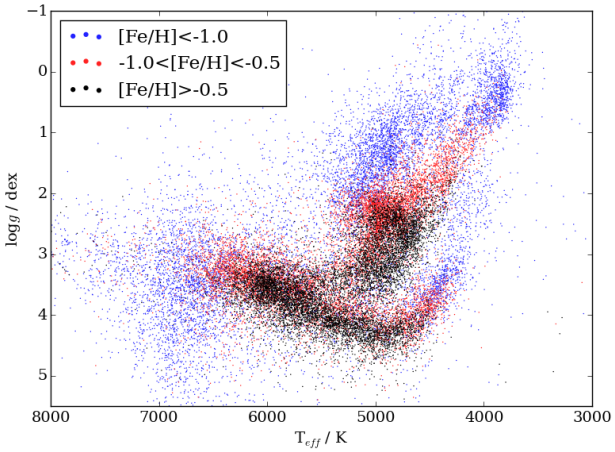

Figure 15. The spectrometric H-R diagram of 10000 randomly selected GALAH stars with metallicity binned into three bins.

neighbours. The final parameters are weighted averages of the ten neighbouring model parameters, with the weights being the normalised linear combinations solution.

Figure 15 shows the HR diagram of 10000 GALAH stars with parameters derived from this method. We note the erroneous upturn at the beginning of the main sequence, a common issue encountered by global fitting algorithms caused by the breakdown of the one dimensional local thermal equilibrium models in cool dwarfs.

7 CONCLUSIONS

In this paper we present the current state of the Iraf reduction pipeline for the GALAH survey. Because the survey is still young and is planed to produce observations for years to come, this will not be the final revision of the code. The observing strategy is expected to remain the same, but there is a possibility that new observing projects will be incorpo-

rated into the GALAH workflow. In the near future maintenance of the HERMES is planned. This will fix the problems with out of focus red arm ccd. Source of the vertical strikes will also be looked for and possibly eliminated. One of the hardest problems to solve when writing this pipeline was creation of linelists used for the wavelength calibration. We are aware of the drawbacks of the ThXe lamp used. At the time of writing first steps were made to mount a laser comb (Steinmetz et al. 2008) onto the 2df. A laser comb is expected to produce much better wavelength calibration in the future, resolution measurements and optical errors corrections. But even with the today's setup, the wavelength calibration is achieved with an accuracy below 1 km/s which is the requirement for the successful determination of elemental abundances, the primary goal of the analysis of stellar spectra in the GALAH project.

We have presented a detailed description of a software pipeline that produces high quality spectra for scientific analysis. There are several analysis pipelines run at different nodes in Australia and Europe. All the analysis pipelines use the same reduced data.

GALAH is the first project to utilize the new spectrograph. Measured behaviour of the HERMES spectrograph in a real-life application will also serve to future users and surveys.

ACKNOWLEDGEMENTS

The data in this paper were based on observations obtained at the Australian Astronomical Observatory as part of the GALAH survey (AAO Programs 2013B/13, 2014A/25, 2015A/19).

JK is funded by ARC grant DP150104667 awarded to JBH and TB.

This research has been supported in part by the Australian Research Council (ARC) funding schemes (grant numbers DP1095368, DP120101815, DP120101237, DP120104562, FS110200035 and FL110100012). SLM acknowledges the support of the ARC through grant DE140100598. DMN was supported by the ARC grant FL110100012.

REFERENCES

AAO software Team 2015, 2dfdr: Data reduction software, Astrophysics Source Code Library (ascl:1505.015)
Ahmed M., Baig M. A., Suleman B., 1998, Journal of Physics B: Atomic, Molecular and Optical Physics, 31, 4017
Blecha A., Cayatte V., North P., Royer F., Simond G., 2000, in Iye M., Moorwood A. F., eds, Proc. SPIEVol. 4008, Optical and IR Telescope Instrumentation and Detectors. pp 467–474
Czesla S., 2015, PyAstronomy, https://github.com/sczesla/PyAstronomy
De Silva G. M., et al., 2015, MNRAS, 449, 2604
Dominquez Palmero L., Jackson R., Molaeinezhad A., FariÃśa C., Balcells M., Benn C. R., 2014, Multi-object spectroscopy data reduction: the AF2+WYFFOS pipeline, doi:10.1117/12.2065314, http://dx.doi.org/10.1117/12.2065314
Geller A. M., Latham D. W., Mathieu R. D., 2015, AJ, 150, 97
Hansen J. E., Persson W., 1987, Physica Scripta, 36, 602
Harris W. E., 1996, AJ, 112, 1487
Humphreys C. J., 1939, J. Res. Natl. Bur. Stand., 22, 19
Humphreys C. J., Meggers W. F., 1933, J. Res. Natl. Bur. Stand., 10, 139
Kausch W., et al., 2015, A&A, 576, A78
Landsman W., 2015, The IDL Astronomy User's Library, https://github.com/wlandsman/IDLAstro
Lovis C., Pepe F., 2007, A&A, 468, 1115
Luo A.-L., et al., 2015, Research in Astronomy and Astrophysics, 15, 1095
Martell S., the GALAH collaboration 2016, The GALAH Survey: Observational Overview, In preparation
Meggers W. F., Humphreys C. J., 1934, J. Res. Natl. Bur. Stand., 13, 293
Nidever D. L., et al., 2015, AJ, 150, 173
Palmer B. A., R. Engleman J., 1993, Technical report, Atlas of the Thorium Spectrum. Los Alamos National Lab, Los Alamos, NM, USA
Redman S. L., Nave G., Sansonetti C. J., 2014, The Astrophysical Journal Supplement Series, 211, 4
Sharma S., the GALAH collaboration 2016, The GALAH Survey: Target Selection and Observation Manager, In preparation
Sheinis A., et al., 2014, in Ground-based and Airborne Instrumentation for Astronomy V. p. 91470Y, doi:10.1117/12.2055595
Simpson J. D., the GALAH team 2016, The GALAH Survey: Relative throughputs of the optical fibres of the 2dF fibre positioner and the HERMES spectrograph, Submitted
Smette, A. et al., 2015, A&A, 576, A77
Steinmetz T., et al., 2008, Science, 321, 1335
Tody D., 1986, in Crawford D. L., ed., Proc. SPIEVol. 627, Instrumentation in astronomy VI. p. 733
Zalubas R., 1976, J. Res. Natl. Bur. Stand., 80A, 221
Zalubas R., Corliss C. H., 1974, J. Res. Natl. Bur. Stand., 78A, 163
de Laverny P., Recio-Blanco A., Worley C. C., Plez B., 2012, A&A, 544, A126
van Dokkum P. G., 2001, PASP, 113, 1420

APPENDIX A: CONVERSION BETWEEN APERTURE, FIBRE AND PIVOT NUMBERS

In this text we mostly use aperture numbers to number the spectra in each image. This is convenient for the reduction process, because nothing has to be done with the spectra of fiducial fibres, as they do not lead any light into the spectrograph, so we only have to deal with real spectral traces. Because the reduction process is the same for all object types, there is no need to know which star actually produces the spectrum we are reducing. The final reduced spectra, however, have to be matched to the correct object. In the observational log each object is assigned a fibre that is numbered by a pivot number on the telescope side and by a fibre number on the spectrograph side. Here we provide a table as a tool to switch between aperture, fibre and pivot numbers. The numbering is different for each plate in some cases.

apt	fib	piv	apt	fib	piv	apt	fib	piv	apt	fib	piv	apt	fib	piv	apt	fib	piv	apt	fib	piv
1	1	10	60	61	70	119	121	130	178	181	190		241	250	295	301	310	354	361	370
2	2	9	61	62	69	120	122	129	179	182	189	237	242	249	296	302	309	355	362	369
3	3	8	62	63	68	121	123	128	180	183	188	238	243	248	297	303	308	356	363	368
4	4	7	63	64	67	122	124	127	181	184	187	239	244	247	298	304	307	357	364	367
5	5	6	64	65	66	123	125	126	182	185	186	240	245	246	299	305	306	358	365	366
6	6	5	65	66	65	124	126	125	183	186	185	241	246	245	300	306	305	359	366	365
7	7	4	66	67	64	125	127	124	184	187	184	242	247	244	301	307	304	360	367	364
8	8	3	67	68	63	126	128	123	185	188	183	243	248	243	302	308	303	361	368	363
9	9	2	68	69	62	127	129	122	186	189	182	244	249	242	303	309	302	362	369	362
10	10	1	69	70	61	128	130	121	187	190	181	245	250	241	304	310	301	363	370	361
11	11	20	70	71	80	129	131	140		191	200	246	251	260	305	311	320	364	371	380
12	12	19	71	72	79	130	132	139	188	192	199	247	252	259	306	312	319	365	372	379
13	13	18	72	73	78	131	133	138	189	193	198	248	253	258	307	313	318	366	373	378
14	14	17	73	74	77	132	134	137	190	194	197	249	254	257	308	314	317	367	374	377
15	15	16	74	75	76	133	135	136	191	195	196	250	255	256	309	315	315/316	368	375	376
16	16	15	75	76	75	134	136	135	192	196	195	251	256	255	310	316	316/315	369	376	375
17	17	14	76	77	74	135	137	134	193	197	194	252	257	254	311	317	314	370	377	374
18	18	13	77	78	73	136	138	133	194	198	193	253	258	253	312	318	313	371	378	373
19	19	12	78	79	72	137	139	132	195	199	192	254	259	252	313	319	312	372	379	372
20	20	11	79	80	71	138	140	131	196	200	191	255	260	251	314	320	311	373	380	371
21	21	30	80	81	90		141	150	197	201	210	256	261	270	315	321	330	374	381	390
22	22	29	81	82	89	139	142	149	198	202	209	257	262	269	316	322	329	375	382	389
23	23	28	82	83	88	140	143	148	199	203	208	258	263	268	317	323	328	376	383	388
24	24	27	83	84	87	141	144	147	200	204	207	259	264	267	318	324	327	377	384	387
25	25	26	84	85	86	142	145	146	201	205	206	260	265	266	319	325	326	378	385	386
26	26	25	85	86	85	143	146	145	202	206	205	261	266	265	320	326	325	379	386	385
27	27	24	86	87	84	144	147	144	203	207	204	262	267	264	321	327	324	380	387	384
28	28	23	87	88	83	145	148	143	204	208	203	263	268	263	322	328	323	381	388	383
29	29	22	88	89	82	146	149	142	205	209	202	264	269	262	323	329	322	382	389	382
30	30	21	89	90	81	147	150	141	206	210	201	265	270	261	324	330	321	383	390	381
31	31	40		91	100	148	151	160	207	211	220	266	271	280	325	331	340		391	400
32	32	39	90	92	99	149	152	159	208	212	219	267	272	279	326	332	339	384	392	399
33	33	38	91	93	98	150	153	158	209	213	218	268	273	278	327	333	338	385	393	398
34	34	37	92	94	97	151	154	157	210	214	217	269	274	277	328	334	337	386	394	397
35	35	36	93	95	96	152	155	156	211	215	216	270	275	276	329	335	336	387	395	396
36	36	35	94	96	95	153	156	155	212	216	215	271	276	275	330	336	335	388	396	395
37	37	34	95	97	94	154	157	154	213	217	214	272	277	274	331	337	334	389	397	394
38	38	33	96	98	93	155	158	153	214	218	213	273	278	273	332	338	333	390	398	393
39	39	32	97	99	92	156	159	152	215	219	212	274	279	272	333	339	332/331	391	399	392
40	40	31	98	100	91	157	160	151	216	220	211	275	280	271	334	340	331/332	392	400	391
	41	50	99	101	110	158	161	170	217	221	230	276	281	290		341	350			
41	42	49	100	102	109	159	162	169	218	222	229	277	282	289	335	342	349			
42	43	48	101	103	108	160	163	168	219	223	228	278	283	288	336	343	348			
43	44	47	102	104	107	161	164	167	220	224	227	279	284	287	337	344	347			
44	45	46	103	105	106	162	165	166	221	225	226	280	285	286	338	345	346			
45	46	45	104	106	105	163	166	165	222	226	225	281	286	285	339	346	345			
46	47	44	105	107	104	164	167	164	223	227	224	282	287	284	340	347	344			
47	48	43	106	108	103	165	168	163	224	228	223	283	288	283	341	348	343			
48	49	42	107	109	102	166	169	162	225	229	222	284	289	282	342	349	342			
49	50	41	108	110	101	167	170	161	226	230	221	285	290	281	343	350	341			
50	51	60	109	111	120	168	171	180	227	231	240		291	299/300	344	351	360			
51	52	59	110	112	119	169	172	179	228	232	239	286	292	298/299	345	352	359			
52	53	58	111	113	118	170	173	178	229	233	238	287	293	297/298	346	353	358			
53	54	57	112	114	117	171	174	177	230	234	237	288	294	300/297	347	354	357			
54	55	56	113	115	116	172	175	176	231	235	236	289	295	296	348	355	355/356			
55	56	55	114	116	115	173	176	175	232	236	235	290	296	295	349	356	356/355			
56	57	54	115	117	114	174	177	174	233	237	234	291	297	294	350	357	354			
57	58	53	116	118	113	175	178	173	234	238	233	292	298	293	351	358	353			
58	59	52	117	119	112	176	179	172	235	239	232	293	299	292	352	359	352			
59	60	51	118	120	111	177	180	171	236	240	231	294	300	291	353	360	351			

Table A1. Conversion between the aperture numbers (apt), fibre numbers (fib) and pivot numbers (piv). There are 400 fibres and pivots, but only 392 apertures in each image. Some fibres and pivots therefore do not have an aperture associated with them. Wherever two pivot numbers are given, the first one is for plate 0 and the second one for plate 1.

APPENDIX B: TH-XE LINELISTS

The tables below give the wavelengths of Thorium and Xe lines used in our linelists. Eventhough 5 decimal places are given for each wavelength we do not claim such accuracy. See section 4.7.1 for the estimation of the precision of the method we used to get these wavelengths. All the lines but one are matched to the lines in the literature, although the match is sometimes ambiguous. Figure B1 shows a typical arc spectrum with matched arc lines for all four arms.

Blue arm

Our λ	Literature λ	Line	Source	Note
4715.03863	4715.18	Xe II	a	†
4723.44378	4723.4382	Th I	b	
4731.12127	4731.1955	Th I	b	
4734.15242	4734.152	Xe I	c	
4740.52510	4740.5292	Th II	b	
4752.40323	4752.4141	Th II	b	
4764.34652	4764.3463	Th I	b	
4769.05694	4769.05	Xe II	a	
4773.23634	4773.241	Th I	b	
4775.21822	4775.18	Xe II	a	
4778.30990	4778.294	Th I	b	
4779.15157	4779.18	Xe II	a	
4784.57700	4784.5667	Th II	d	
4786.51691	4786.531	Th I	b	
4787.76840	4787.7462	Th II	d	♣
4789.44399		Th I, Th II	e, b	♣♣
4792.61260	4792.619	Xe I	c	
4796.40438	4796.48	Xe II	a	
4799.32645	4799.45	Xe II	a	∗
4807.02831	4807.02	Xe I	c	
4808.14795	4808.1337	Th I	b	
4809.60832	4809.614	Th I	b	
4817.02113	4817.0206	Th I	b	
4818.00478	4818.02	Xe II	a	
4823.27491	4823.35	Xe II	a	∗∗
4826.67568	4826.7004	Th I	b	
4829.69964	4829.71	Xe I	c	
4831.13337	4831.1213	Th I	b	
4840.88300	4840.8492	Th I	b	
4843.28996	4843.29	Xe I	c	
4844.32227	4844.33	Xe II	a	
4848.34116	4848.3625	Th I	b	
4850.39484	4850.23	Xe II	a	∗
4853.74521	4853.77	Xe II	a	
4862.35682	4862.45	Xe II	a	∗
4863.16331	4863.1724	Th I	b	
4865.47257	4865.4775	Th I	b	
4872.91990	4872.9169	Th I	b	
4876.46465	4876.4949	Th I	b	
4878.74379	4878.733	Th I	b	
4883.52998	4883.53	Xe II	a	
4884.07236	4884.07	Xe II	a	∗∗
4887.28188	4887.29	Xe II	a	
4890.05617	4890.09	Xe II	a	
4894.95091	4894.9551	Th I	b	
4902.04918	4902.0545	Th I	b	

Table B1. Lines from our blue arm linelist and wavelengths from the literature. First column gives the wavelengths we assumed for our linelist. Second column gives wavelengths of the same lines as found in the literature. Third column shows the element and its ionization state responsible for the line. Literature source and notes are in the remaining two columns. Literature sources: a: Hansen & Persson (1987), b: Palmer & R. Engleman (1993), c: Humphreys & Meggers (1933), d: Lovis & Pepe (2007), e: Redman et al. (2014), f: Humphreys (1939), g: Meggers & Humphreys (1934), h: Zalubas (1976), j: Ahmed et al. (1998), k: Zalubas & Corliss (1974). Notes: †: Wavelengths don't match, but there is a line nearby in the literature. Hence we put the wavelength of the nearest line into the table. ∗: Line is marked hazy in the literature. ∗∗: line is marked very hazy in the literature. ‡: The wavelength does not match, but there are two lines nearby, one of Xe II and one of Th I. ♣: This is a blend of two equally strong lines. The average wavelength is given in the table. ♣♣: This is probably a blend of two Thorium lines. They have different strength in the literature and come from different ionization stages. It is therefore impossible to calculate a precise average wavelength.

Green arm

Our λ	Literature λ	Line	Source	Note
5650.28289	5650.3276	Th I	b	
5657.94536	5657.9255	Th I	b	
5659.39468	5659.38	Xe II	f	
5665.21802	5665.1799	Th I	b	
5667.55524	5667.56	Xe II	f	
5670.85448	5670.91	Xe II	f	
5685.22615	5685.1921	Th I	b	
5695.81028	5695.75	Xe I	c	
5696.54602	5696.477	Xe I	g	
5699.62286	5699.61	Xe II	f	
5700.96596	5700.9176	Th II	b	
5707.11348	5707.1033	Th II	b	
5715.98484	5716.10	Xe II	f	
5719.62383	5719.61	Xe II	a	
5720.20852	5720.1828	Th I	b	
5725.40398	5725.3885	Th I	b	
5726.91340	5726.91	Xe II	f	
5741.88955	5741.82881	Th I	e	
5744.12509	5744.2	Xe II	a	
5748.76371	5748.7412	Th I	b	
5749.37855	5749.3883	Th II	b	
5751.01375	5751.03	Xe II	f	
5753.08690	5753.0265	Th I	b	
5758.67208	5758.65	Xe II	f	
5760.57706	5760.5508	Th I	b	
5763.57841	5763.529	Th I	b	
5768.14854	5768.1812	Th I	b	
5776.39142	5776.39	Xe II	f	
5789.65389	5789.6451	Th I	b	
5792.44789	5792.4304	Th I	b	
5797.47803	5797.3194	Th I	b	†
5800.84347	5800.8297	Th I	b	
5804.17143	5804.1412	Th I	b	
5814.51830	5814.505	Xe I	g	
5815.34288	5815.4219	Th II	b	†
5815.92798	5815.96	Xe II	f	
5823.88112	5823.89	Xe I	c	
5824.76730	5824.80	Xe I	c	
5834.99388	5835.167	Th I	d	†
5856.50900	5856.509	Xe I	g	
5859.37251	5859.348	Th I	h	
5868.63003	5868.37475	Th I	e	†
5875.01800	5875.02	Xe I	c	

Table B1. Table continues for the green arm.

Red arm

Our λ	Literature λ	Line	Source	Note
6485.62436	6485.3759	Th II	b	†
6487.76893	6487.76	Xe I	c	
6490.76162	6490.7372	Th I	b	
6494.76087	6494.981	Th II	d	†
6498.71437	6498.72	Xe I	c	
6504.22250	6504.18	Xe I	c	
6506.93387	6506.9863	Th I	b	
6509.10027	6509.0503	Th I	b	
6512.78374	6512.83	Xe II	f	
6521.48606	6521.508	Xe I	g	
6528.61811	6528.65	Xe II	f	
6531.35671	6531.3418	Th I	b	
6533.16006	6533.16	Xe I	c	
6543.34982	6543.360	Xe I	g	
6549.35230	6549.820	Th I	h	†
6551.75606	6551.7055	Th I	b	
6552.71340	6552.79	Xe I	j	♣
6554.20081	6554.196	Xe I	g	
6556.75037	6556.70	Xe II	a	
6560.04011	6559.97	Xe I	c	
6563.17049	6563.19	Xe II	a	
6577.22645	6577.2146	Th I	b	
6583.88861	6583.906	Th I	b	
6588.55275	6588.5396	Th I	b	
6591.48314	6591.4845	Th I	b	
6593.95197	6593.9391	Th I	b	
6594.97730	6595.01	Xe II	f	
6595.56355	6595.56	Xe I	c	
6597.23314	6597.25	Xe II	f	
6598.81354	6598.84	Xe II	f	
6602.75824	6602.7620	Th I	b	
6619.98306	6620.02	Xe II	a	
6632.42826	6632.464	Xe I	g	
6644.65341	6644.6700	Th I	b	
6657.88712	6657.92	Xe I	c	
6662.25522	6662.2686	Th I	b	
6666.94783	6666.965	Xe I	g	
6668.90004	6668.92	Xe I	c	
6674.65633	6674.6969	Th I	b	
6694.25904	6694.32	Xe II	f	
6694.92523	6695.181	Th I	d	†
6702.19461	6702.25	Xe II	a	
6703.85395	6703.8966	Th II	b	
6719.17725	6719.19943	Th I	e	
6728.00680	6728.01	Xe I	c	

Table B1. Table continues for the red arm.

IR arm

Our λ	Literature λ	Line	Source	Note
7600.76466	7600.77	Xe I	c	
7618.51401	7618.57	Xe II	f	
7627.20055	7627.1749	Th I	b	
7630.26763	7630.3106	Th I	b	
7636.21145	7636.17473	Th I	e	
7642.00652	7642.02	Xe I	c	
7643.92543	7643.91	Xe I	c	
7647.40741	7647.3794	Th I	b	
7653.84500	7653.8284	Th I	b	
7658.29378	7658.3202	Th I	b	
7664.58740	7664.56	Xe I	c	
7666.66991	7666.61	Xe I	c	
7670.62263	7670.66	Xe II	f	
7674.41459	N/A	N/A	N/A	
7678.12287	7678.1267	Th I	b	
7685.30565	7685.3075	Th II	b	
7687.74823	N/A	N/A	N/A	
7698.97833	N/A	N/A	N/A	
7710.24928	7710.2689	Th I	b	
7712.24919	7712.42,7712.405	Th I, Xe II	a	‡
7728.94825	7728.9509	Th I	b	
7740.35102	7740.31	Xe I	c	
7783.66085	7783.66	Xe I	c	
7786.96975	7787.04	Xe II	f	
7788.94932	7788.9342	Th I	b	
7798.38957	7798.35789	Th I	e	
7802.65468	7802.65	Xe I	c	
7817.70870	7817.7669	Th I	b	
7832.97925	7832.98	Xe I	c	
7837.00751	7836.703	Th II	k	†
7847.50351	7847.5394	Th I	b	
7865.97327	7865.9698	Th I	b	
7881.27733	7881.32	Xe I	c	
7886.27084	7886.283	Th I	b	

Table B1. Table continues for the IR arm.

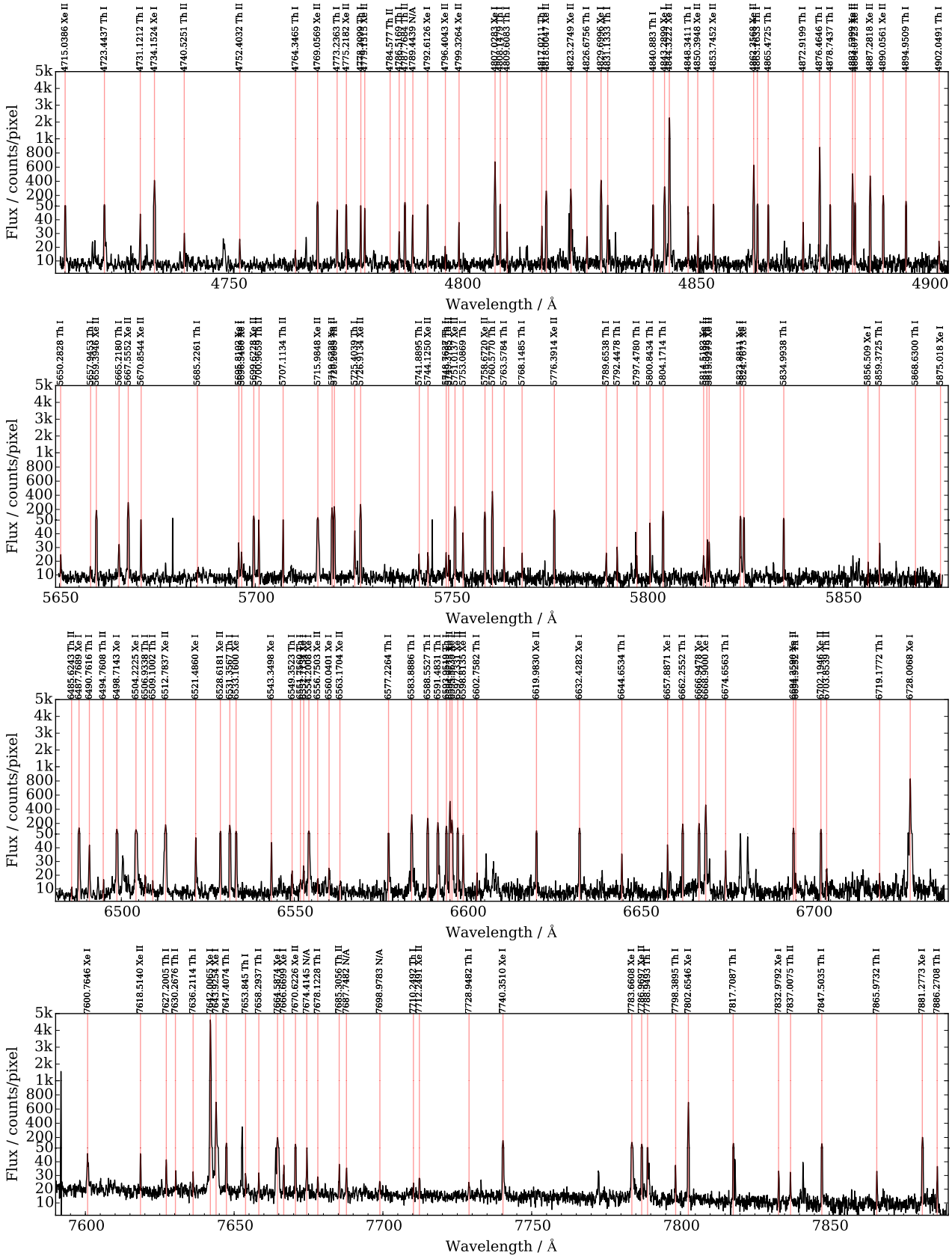

Figure B1. Typical arc spectra with marked lines included in the linelists. Blue, green, red and IR arms follow top to bottom. Some strong, but unmarked lines are cosmetic errors and some unmarked lines have peculiar shapes, so the did not make it into the linelists. The wavelengths are taken from our linelist. There is a lot of dynamic range in the flux, so we used a nonlinear flux scale in all four panels.

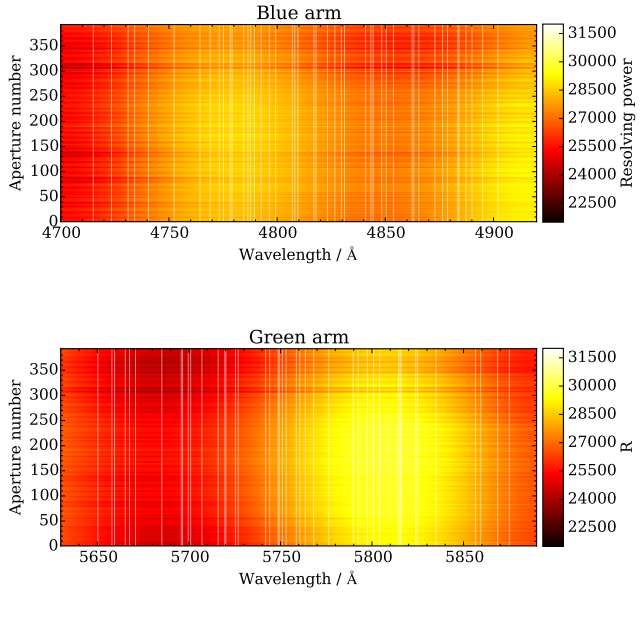

Figure C1. Resolution maps for the blue arm (top) and the green arm (bottom). The maps show the colour coded resolving power as it depends on the wavelength and aperture number. White vertical lines show arc lines that were used to produce these maps. Map is composed of 6th degree polynomials that represent the resolving power for each aperture. Polynomials were fitted to measured arc lines widths for each aperture independently.

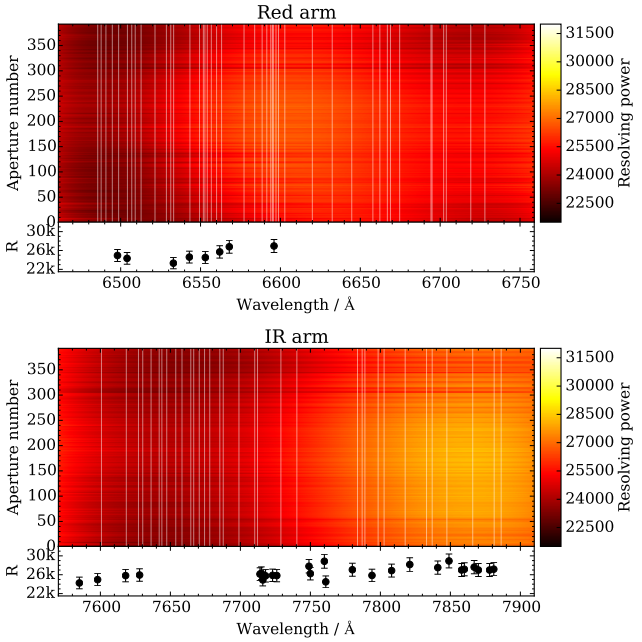

Figure C2. Resolution maps for the red arm (top) and the IR arm (bottom). The maps show the colour coded resolving power as it depends on the wavelength and aperture number. White vertical lines show arc lines that were used to produce these maps. Map is composed of 6th degree polynomials that represent the resolving power for each aperture. Polynomials were fitted to measured arc lines widths for each aperture independently. The smaller panels below the resolution map show how the resolving power changes with wavelength when measured from sky emission lines. Each dot represents the resolving power measured from one line and the errorbars show the scatter accumulated over many images. There are no detectable sky emission lines in the blue and green arm. The resolution in the case of sky emission lines was measured from sky spectra combined from all the sky fibres in the image, so no dependence on aperture number is given.

APPENDIX C: RESOLUTION MAPS

Light and later signal go through many optical elements and computational processes, so it is not trivial to calculate the resolution of final reduced spectra. It is however critical that the real resolution is known by analysis methods. We therefore calculated the resolution from observed arc spectra. Arc spectra consist of unresolved emission lines, conveniently distributed over most of the observed wavelengths. Arc lines are strong enough to have their profile measured in individual spectra. The alternative are sky emission lines, also unresolved but weaker than arc lines and found only in parts of the red and IR arm.

We calculated the full width at half maximum (FWHM) for each arc line in all the allocated fibres for over 200 images randomly selected from the whole survey campaign. The resolution maps show the resolving power:

$$R = \frac{\lambda}{FWHM}, \tag{C1}$$

which was calculated for each measured line. A 6th degree polynomial was fitted to all the measurements in each aperture. Because some arc lines are blends of two unresolved lines, a rejection algorithm was used to fit the polynomial only to measurements of single lines. A polynomial was fitted independently for each aperture and the result is plotted in Figures C1 and C2.

The emerged pattern was not expected, so we tried to verify the waviness along the wavelength axis by measuring the resolution from sky emission lines. Because sky emission lines are weak in individual spectra, we used combined sky spectra, like the one in Figure 10. The result is shown in Figure C1 under the red and IR maps. There are no obvious sky emission lines in the green and blue arms and only a few in the red arm, so the comparison is humble. We observe a correlation in the resolution measured in both ways, so we believe that the wavy pattern is the consequence of the instrument and not of the inconveniently blended arc lines.

The measured resolution is in a rough agreement with the resolving power of 28000 for which the instrument was designed. Exception is the red arm, where the lower average resolution is due to out of focus CCD, a mechanical error that will be fixed during the planned maintenance of the HERMES spectrograph in late 2016.

This paper has been typeset from a TEX/LATEX file prepared by the author.