Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **464,** 1259–1281 (2017)
Advance Access publication 2016 August 17

doi:10.1093/mnras/stw2064

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] Andrew R. Casey,[4] Gayandhi M. De Silva,[5,1] Ken C. Freeman,[3] Sarah L. Martell,[6] Jeffrey D. Simpson,[5] Katharine J. Schlesinger,[3] Daniel Zucker,[7,8] Borja Anguiano,[7] Carlos Bacigalupo,[7] Timothy R. Bedding,[1,9] Christopher Betters,[1,10] Gary Da Costa,[3] Ly Duong,[3] Elaina Hyde,[11] Michael Ireland,[3] Prajwal R. Kafle,[12] Sergio Leon-Saval,[1,10] Geraint F. Lewis,[1] Ulisse Munari,[13] David Nataf,[3] Dennis Stello,[1,9] C. G. Tinney,[6,14] Gregor Traven,[2] Fred Watson[5] and Robert A. Wittenmyer[6,14]

Affiliations are listed at the end of the paper

Accepted 2016 August 16. Received 2016 August 15; in original form 2016 May 11

ABSTRACT

We present the data reduction procedures being used by the GALactic Archeology with Hermes (GALAH) survey, carried out with the HERMES fibre-fed, multi-object spectrograph on the 3.9-m Anglo-Australian Telescope. GALAH is a unique survey, targeting 1 million stars brighter than magnitude $V = 14$ at a resolution of 28 000 with a goal to measure the abundances of 29 elements. Such a large number of high-resolution spectra necessitate the development of a reduction pipeline optimized for speed, accuracy, and consistency. We outline the design and structure of the IRAF-based reduction pipeline that we developed, specifically for GALAH, to produce fully calibrated spectra aimed for subsequent stellar atmospheric parameter estimation. The pipeline takes advantage of existing IRAF routines and other readily available software so as to be simple to maintain, testable, and reliable. A radial velocity and stellar atmospheric parameter estimator code is also presented, which is used for further data analysis and yields a useful verification of the reduction quality. We have used this estimator to quantify the data quality of GALAH for fibre cross-talk level ($\lesssim$0.5 per cent) and scattered light ($\sim$5 counts in a typical 20 min exposure), resolution across the field, sky spectrum properties, wavelength solution reliability (better than 1 km s^{-1} accuracy), and radial velocity precision.

Key words: atmospheric effects – instrumentation: spectrographs – methods: observational – techniques: spectroscopic – surveys – stars: atmospheres.

1 INTRODUCTION

There is no definitive recipe for reducing multifibre spectra. Multifibre, multi-objects spectrographs are sophisticated and complex instruments. The complexity of these instruments precludes the analysis of its data by individual users on their own, so the users rely on reduction pipelines provided by the instrument team. To date each such instrument (e.g. Blecha et al. 2000; Dominquez Palmero et al. 2014; Luo et al. 2015; Nidever et al. 2015) has developed a dedicated data-processing pipeline that delivers processed spectra for that instrument to its users (including large surveys being carried out with the instrument). The quality of any scientific analysis depends strongly on the quality of the reduced spectrum. It is critical that users understand the details of the reduction pipeline, such as the reduction process, steps made, and respective levels of calibration and the accuracy of the delivered data to take a full advantage of the products.

The GALAH (GALactic Archeology with Hermes) survey (De Silva et al. 2015) uses the HERMES fibre-fed, multi-object spectrograph at 3.9-m Anglo-Australian Telescope (AAT) – itself fed by the 2dF fibre positioner system. The spectrograph delivers high-resolution spectra in four wavelength ranges. By combining multi-object capability with a high-resolution spectrograph, the GALAH survey aims to measure the abundances of 29 elements for one million stars over a 5 yr period. GALAH is the first major survey being

★ E-mail: janez.kos@sydney.edu.au

© 2016 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

conducted with this spectrograph which was designed with survey goals in mind (Sheinis et al. 2015). The fibre positioner system (2dF), however, is older and used by several other multifibre spectrographs – primarily for surveys targeting extragalactic sources (e.g. Colless et al. 2003; Croom et al. 2004; Drinkwater et al. 2010). These have all made use of a common reduction pipeline called 2dfdr (AAO software Team 2015). 2dfdr serves several different spectrographs that all use the common 2dF fibre positioner system. A reduction pipeline within the 2dfdr environment was also written for HERMES, but it was not suitably optimized for GALAH (unreliable wavelength calibration) and lacked many important features (sophisticated sky removal algorithm), so it often produced results of poor quality. Implementation of features 2dfdr lacks is easy in IRAF, as it is more general purpose software and gives the user options that have not been considered yet in 2dfdr.

We have therefore developed an alternative pipeline for the GALAH survey. It is based on pre-existing IRAF[1] routines in order to be simple to maintain. IRAF is convenient to use, because more time and care can be invested into building functional and advanced routines that constitute a pipeline instead of writing the functions themselves that already exist in the IRAF distribution. IRAF functions, however, are not enough to create a pipeline. They serve a very general purpose and we use it as any library of computer code in the process described in this paper. IRAF has been widely used in the astronomy community for more than two decades. The routines we use have therefore been extensively tested and proved to be reliable (Tody 1986). Our IRAF pipeline is now the default pipeline used in the GALAH survey. While it is regularly updated and improved, it remains an internal project, and is not intended for general HERMES users. Some results and analysis of the HERMES spectrograph, on the other hand, serve a much broader audience and are not specific to the GALAH survey. The same holds for some routines we developed [like cosmic removal from images with a strong pattern, fitting and removing sky spectra, and calibration of the Thorium–Xenon (ThXe) arc lamp] that serve all HERMES users and other spectrographs that are dealing with the same problems.

This paper briefly reviews observational aspects of the GALAH survey in Section 2 and the motivation for a new reduction pipeline in Section 3. The main part of the paper is a step-by-step description of the pipeline (Section 4). In Section 5, we describe the organization of reduced data. As a part of the reduction pipeline we also run another code, a simple estimator of the radial velocity, and basic parameters of stellar atmospheres. This is discussed in Section 6. Two approaches to measure the resolution of the GALAH spectra are presented in Section 7. An important by-product of this work – the line lists for the wavelength-calibration lamp are collected in the appendix.

2 OBSERVING PROCEDURE

The observing procedures for the GALAH survey are described in detail elsewhere (Martell et al., in preparation; Sharma et al., in preparation), so we only summarize relevant parts here.

HERMES is a high-resolution, multifibre spectrograph, simultaneously observing four wavelength ranges in different arms (see Table 1). The resolving power is 28 000 in normal observing mode (used for the GALAH survey) or 45 000 in a slit-mask mode. The resolution in the GALAH observing mode corresponds to a resolution element size of 3.5 pixels on each of the four CCDs. Note that due to the design of the spectrograph, the wavelength ranges vary by a few Ångstroms with the position of the spectrum on the CCD.

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

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

The GALAH survey fields are defined by a fixed pattern across the southern sky (Sharma et al., in preparation). Each field has a unique observing configuration file with information for each star within its sky area. 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 the optimal fibre arrangement to observe as many targets in the field as possible. It also allocates sky fibres and selects guide stars. Once the observer is satisfied with the configuration, it is sent to the 2dF fibre positioner robot, which re-arranges the fibres into the correct positions.

Fibres are held in position with magnets on one of two metal plates (called plate 0 and plate 1). These face away from each other, such that 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 each plate available to the 2dF positioner robot. Of these, 392 are regular fibres and 8 are fiducial (auxiliary) fibres used for guiding. On the telescope end, fibres on each plate are given a so-called pivot number from 1 to 400. Bundles of 10 fibres are then fed into the spectrograph where they form a pseudo-slit. Each such group of 10 fibres is called a slitlet. In each slitlet, there is an optical element that changes the focal ratio of the fibres. This inverts 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, fibre numbers do not correspond directly to pivot numbers. During the reduction, we ignore the fiducial fibres (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 again numbered (this time from 1 to 392), so these aperture numbers do not correspond to either fibre or pivot numbers. Conversion between pivot and fibre numbers is given in the table A1 in Simpson et al. (2016). A further conversion between fibre and aperture numbers is simple: the aperture number follows the same order as the fibre number, but skips eight fiducial fibres. Fire number one is therefore aperture number one, but fibre number 400 is aperture number 392.

At the beginning of each night, observers take a series of bias frames for each CCD. These are used for the reduction of the science, flat and arc lamp images and do not have to be repeated later in the night. At the same time, the first two fields are sent to the fibre configuration software and the fibres are positioned on both plates. After the first field is observed, the plates are switched and the second field is observed, while another field is configured on the now available first plate.

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

[1] IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

same fibre configuration as for the science exposures. Not all fields require 1 h of exposure time. During twilight we observe bright fields that require less exposure time, while special programmes (such as observations of standard stars or K2 follow-up) also use different exposure times.

The images created by the four CCDs (one per spectrograph arm) are kept in separate folders and are identified by their run number, which follows the consecutive exposure done in that night. The fibre configuration used (including sky, fiducial, and unused fibres) is also saved – to record which star was positioned in which fibre.

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. However, the software is designed for many different instruments with a wide range of data types. Thus, 2dfdr lacked specific functionality required to cope 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 adjustments can be implemented quickly. The pipeline is fine-tuned to work only with GALAH spectra that have been taken in a consistent 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 have made extensive use of pre-existing and already tested IRAF routines. For procedures not available in the IRAF packages, we, again, used existing solutions before writing our own. Due to its foundation in IRAF, we refer to this pipeline simply as the 'IRAF pipeline'. Because the pipeline uses readily available software, the whole pipeline contains only ∼2000 lines of code. The cost of such approach is a slow execution of the code. The reduction (including radial velocity and stellar atmospheric parameter estimation) of the current GALAH data set (at the time of writing 200 000 unique stars or 600 fields or around 100 nights assuming the maximum efficiency) takes almost 2 weeks on a top-end desktop PC. The main limitation is that IRAF uses many hard drive read/write operations on files at numerous steps. In spite of the speed, IRAF had many advantages that suited our requirements.

We have striven to make the pipeline automatic, requiring no human interaction to reduce the data. However, for some steps, complete automation requires disproportionate effort to implement, so human intervention is needed when observing strategy is changed, or spectrograph properties change.

It was also felt to be desirable that the same code is used to reduce spectra from all four arms to simplify the pipeline significantly, which has been achieved in the pipeline presented here, except for a few steps that are skipped entirely for spectra from some arms because they are unnecessary.

4 THE PIPELINE

The major stages used by the pipeline are presented here in a logical order, but note that they are not necessarily executed in the same order, due to complications like an individual step requiring multiple iterations.

The IRAF pipeline reduces data on a night-by-night basis, and the same procedure is used for all images, regardless of which observing programme produced them. Before submission to the pipeline, images are separated into eight groups – one for each of the four arms in either of the two plates. Each group of images is then reduced on its own without any interaction with the processing for other groups. Images from each night are reduced separately from all other nights. Any number of executions of the pipeline can be run in parallel, making the reduction process faster and more efficient. Usually we run between 16 and 32 jobs at once, so two to four nights can be reduced in parallel.

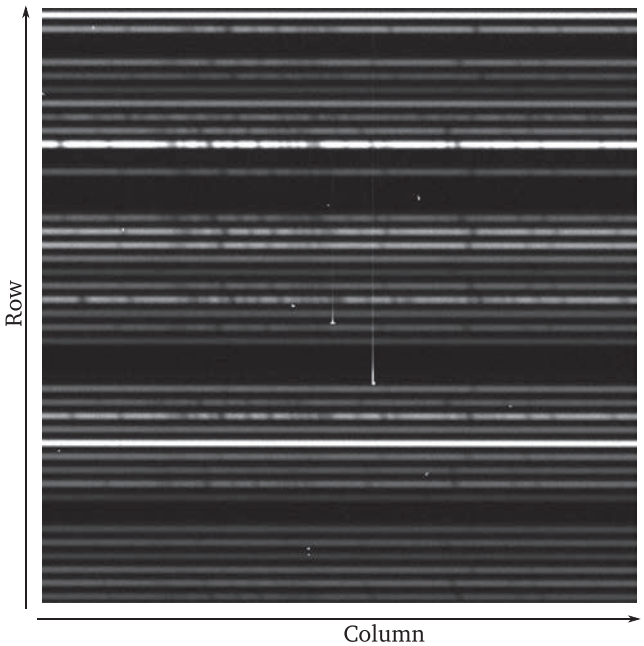

Figure 1. An example of two vertical streaks in a science exposure, made on 2015 February 8, run 39, green arm. Horizontal lines are traces of stellar spectra with clearly visible stellar absorption lines. Two streaks are visible in the centre. A few cosmic rays are also visible. Size of this image excerpt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

4.1 Non-automated routines

As noted earlier, some pipeline steps cannot be completely automated. They do not have to be performed every time the pipeline is run – some are done once for all nights, while others are updated from time to time – usually as a result of a changed instrument calibration or a new observing strategy.

Spectrum traces (see Fig. 1 for an example of how raw HERMES spectra look like) are found by taking a cross-section along a specified CCD column and locating the peaks representing the traces. The trace-finding algorithm relies heavily on a 'key' which indicates which aperture is positioned at which pixel. The position and separation between traces are not uniform and can also change when cameras are refocused or CCDs moved. A key was produced that provides the location of the trace centre for each aperture in a 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. In comparison, the full width at half-maximum (FWHM) of the trace is around 4 pixels in the spatial direction and the aperture width used for spectral extraction is 6 pixels. The key has to be generated by hand and can be used for as long as the geometry of the traces does not change significantly. At the time of writing this has only been done once.

Once the traces are identified in the reference column, they are propagated to other columns with the same criterion that the deviation from the key is of 3 pixels at most. This guarantees that the apertures are positioned on the correct spectra every time and that the trace does not jump between the neighbouring spectra.

Wavelength calibration also requires some human intervention. A lot of time has been invested in building the line lists for the wavelength calibrations (see Section 4.7 for a detailed description). The line lists alone, however, are not enough to make a wavelength calibration. The method also expects to be given an approximate relationship to transform pixels into wavelengths with sufficient precision such that correct arc lines are identified. The wavelength calibration was initially done manually for a large number of spectra, which allowed this relationship to be established. Humans are much more likely to notice mismatched lines than computers and the established relationship is used as a proxy for the wavelength solution. We monitor the final wavelength solution from time to time to determine if the proxy relationship has to be updated for new data.

We use human verification to assist with cosmic ray removal. An important parameter that was found by eye is the threshold in the cosmic ray removal algorithm. If this removal threshold is set too low, useful data will be tagged as cosmic rays, significantly impacting 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 run through the cosmic ray removal algorithm and inspected by eye. The thresholds chosen as a result were then hard-coded to give the optimal result in every type of image we checked.

Finally, a large number of parameters (e.g. orders of fitted polynomials) were also determined by reducing a few images by hand and adjusting the parameters to get the best results. They are explained in their appropriate sections below.

4.2 Cosmetic corrections

With each set of science frames, we take one flat frame and one arc frame, either before or after the science frames. The flat frame is a so-called flap flat, produced by a system of lights shining on flaps inserted into the telescope below the prime focus location of the 2dF, that direct lamp light on to the fibre ends. Not only is a truly flat illumination not guaranteed with such a set-up, but the flat light enters the fibres at different angles from sky illumination and so the amount of light that actually enters each fibre depends on the fibre position within the field. Dome flats (a flat-field produced by reflecting light off a screen mounted on the dome) and sky flats are taken from time to time, but not every night. These two give an illumination of fibres that is much closer to that of stars.

One flat-field is created from all flap flats taken in one night for each arm and each plate. Since the fibres are repositioned roughly every hour, this averages out uneven illumination and produces evenly illuminated and consistent flats to use in the pipeline. See also Section 4.8 for a description of the sky-subtraction process, which is highly reliant on correct flat-field reduction. Success in removing the sky emission lines is proof that averaging several flap flats produces a usable flat-field image.

A flap flat-field does not illuminate the whole CCD uniformly, so we do not correct images for a flat-field response in a conventional way. Dome flats are not taken as part of a regular observing run due to time constraints and the twilight flats constitute the solar spectrum, which creates additional complications in the flat-field correction. Due to simplicity, we therefore use flap flats only everywhere in the pipeline. We also do not use dark frames, as the HERMES CCDs are cooled to levels where dark current is negligible (measured dark current is less than 1 electron per hour). The first step in the reduction process is then bias correction. We use a median of all bias frames (typically there are 15 bias frames taken for each arm) for this, and we check that the bias levels are the same in the overscan of every image as in the combined bias image. If not, the overscan is used and subtracted instead of the combined bias frame.

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

Vertical streaks are suspected to be very energetic particles originating from the HERMES spectrograph itself. The source has not been confirmed yet, though we suspect either one of the optical components or its coatings. There are around 20 such events detected per 20 min exposure in the blue and green arms, only a few in the red arm, and none in the IR arm. The energetic events responsible for the streaks are visible in the IR arm but they do not produce the streak. A vertical strike is different from a traditional cosmic ray – the strike saturates several pixels at the impact point and blooms into pixels in the same column (see Fig. 1). The signal continuing from the blooming part becomes weaker with distance from the impact point and degrades to undetectable levels after 500 pixels in the most extreme cases and around 80 in a typical case. This means that a single vertical streak event impacts many spectra. Some vertical streaks 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 ray removal procedure. What is left is the centre of the streak (if it is large), and the weak tails of the streaks which do not have enough contrast to be detected by a cosmic ray removal algorithm.

Degraded columns are only present in the IR arm on the first analogue-to-digital converter. There are six such columns (columns numbers 3262, 3270, 3271, 3308, 3324, and 3347) and the signal in them is replaced by linearly interpolated flux from nearest unaffected columns with the IRAF's `ccdproc` routine. They are all in the region of the spectrum uncorrupted by the telluric *A* band.

We are using a well-known and well-tested algorithm for detecting and removing cosmic rays called LaCosmic (Van Dokkum 2001). A PYTHON version of the code is being used with minor modifications. The method described by Van Dokkum (2001) was designed to find and remove cosmic rays in images with no apparent patterns or symmetries, such as plain images of astronomical objects. However, images of spectra have a strong pattern in them (i.e. horizontal spectral traces). When a cosmic ray is detected, the flux in affected pixels can no longer be interpolated from the neighbouring pixels without considering this pattern. In the release of the code originally adopted by us, 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 box size takes the obvious spectral pattern into account, and the interpolated value will be calculated from the pixels lying on the same spectral trace. Note that each spectral resolution element is around 4 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 a pixel affected by a cosmic ray is minimal. This approach is not perfect, as the spectra also have a predictable pattern in the axis perpendicular to the spectral traces – a distinct bell

shape. Just changing the box shape, however, introduced a significant improvement.

In the blue and green arms, where we have strong vertical streaks, 0.073 per cent of pixels are detected to be affected by cosmic rays by LaCosmic in a 20 min exposure. This drops to 0.062 per cent in the red and IR arms. We estimate that 99 per cent or more of the pixels affected by cosmic rays and vertical streaks are corrected. We are satisfied with the LaCosmic performance, so this is the only measure made to remove the cosmic rays. Even though we make three consecutive exposures of each field, we do not take their median at any stage. Median combination of, say normalized spectra, is often considered a good step to remove cosmic rays, but in our case it would fail to remove many vertical streaks 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 a whole group of images (i.e. one night, one CCD, and one plate).

392 extraction apertures must be positioned on the spectral traces in each image. A key (as described above) is used to give the pipeline an approximate position for each aperture in one reference column (see Section 4.1) of the image. After the centre of the aperture is calculated at the reference column, it is propagated along the spectrum trace. IRAF's `center1d` algorithm is used to find the aperture centres along the trace. 10 consecutive columns are averaged before the centre of the trace is found, so the algorithm is run on higher SNR data and fewer times for each trace. The shape of the trace is represented by a fifth order Chebyshev polynomial. The traces are almost straight, well behaved, and follow one row of pixels with a deviation of 3 pixels at most (see Fig. 2). The width of the aperture is fixed at 6 pixels and is centred on the fitted trace. Pixelization is taken care of by IRAF's `apfit` and `fit1d` algorithms.

4.4 Correction of optical aberrations

HERMES is designed to produce almost linear, equally spaced, two-dimensional spectral traces. This is achieved to a high degree. Optical aberrations are most prominent in the corners of the CCD,

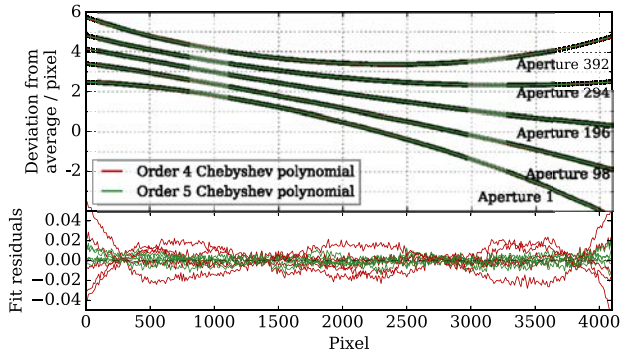

Figure 2. Traces fitted in the apertures 1, 98, 196, 294, and 392 in a green arm flat-field show the geometry of the spectral traces (top panel). Plotted traces are shifted for 1 pixel each in vertical direction. A Chebyshev polynomial of order 4 and 5 is fitted to each trace and the residuals are plotted in the bottom panel.

where the point spread function (PSF) takes an obvious elliptical shape. A similar distortion is observed in all four arms. 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 empirically, by using arc lines to measure the tilt of the elongated PSF and straightening the spectra to narrow the PSF in the wavelength direction (see Fig. 3 for an example of a 'tilted' PSF and the desired correction). This way we can improve the spectral resolution in the detector corners. We conclude that the PSF is adequately sampled by arc lines for this operation. We do this process for every aperture individually.

Each aperture is divided into seven parallel and equal-width sub-apertures and a spectrum is extracted from every sub-aperture. This is performed by IRAF's `apall` routine with option `nsubaps` set to 7. Arc lines are identified in each sub-aperture, but no wavelength calibration is done. Only the positions of centres of the lines in the pixel space are recorded. They are calculated by IRAF's `center1d` algorithm. It is then assumed that the centre of an arc line should have the same coordinate in every sub-aperture, and this gives us a condition for a transformation. For every arc line, we want to transform the coordinates of the line centres in each sub-aperture to match the global coordinate, as measured if there was only one

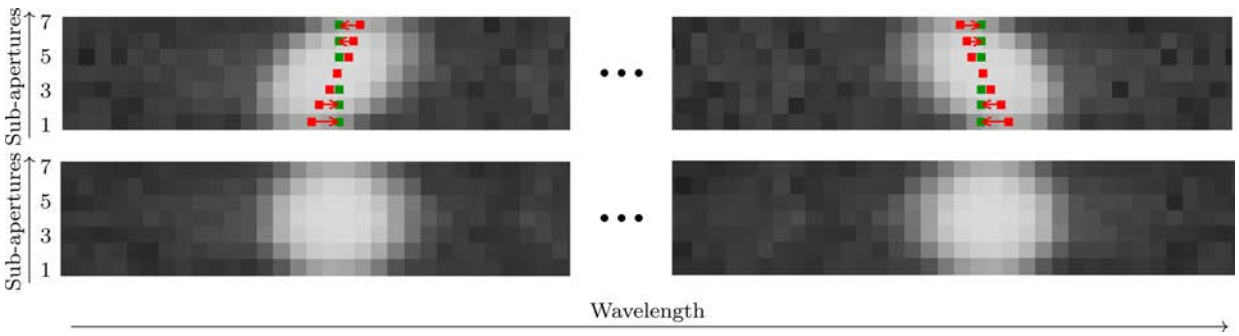

Figure 3. Example of two arc lines in the corners of a red arm image, before (top) and after (bottom) the geometric correction. Left image shows the shape of an arc line in the top left corner and right image shows it in the top right corner of an image. Notice that optical aberrations distort the two arc lines differently. The aperture is divided into seven sub-apertures, seen here as rows of pixels. Sub-apertures are equal division of an aperture and are not centred on pixels but run parallel to the spectrum trace. The calculated centre of the line in each sub-aperture 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 on to the green ones, as illustrated by red arrows. The resolution in the bottom panel is improved by ∼15 per cent over the resolution in the top panel.

aperture:

$$(x_1, 1) \rightarrow (x_f, 1)$$
$$(x_2, 2) \rightarrow (x_f, 2)$$
$$\vdots$$
$$(x_7, 7) \rightarrow (x_f, 7), \quad (1)$$

where x is the coordinate in pixels along the wavelength axis, and x_f is the global coordinate of each arc line. The second coordinate component is just the sub-aperture number (see Figs 3 and 5 for an illustration).

With ∼40 arc lines per spectrum and seven data points coming from the above transformation condition for each line we have hundreds of data points to constrain the following transformation. The transformation adopted is a linear skewness that can change along the wavelength axis according to a third-order polynomial. Following the rule in equation (1), a transformation is written for each pixel:

$$y = \mathrm{S}(x)(x - x_f) \quad (2)$$

where $\mathrm{S}(x) = Ax^3 + Bx^2 + Cx + D$ is a third-order polynomial describing how the shear is changing along the x-axis. What we are interested in is the $x_f(x, y)$, the shift we have to apply to each pixel, which can be calculated from the formula

$$x_f(x, y) = x - \frac{y}{Ax^3 + Bx^2 + Cx + D} \quad (3)$$

after the coefficients A to D are fitted to measured points by minimizing the difference between the x_f from equation (3) and the measured shifts for each line (as shown in equation 1). Five iterations of the fitting process are done, with outliers above 2.5σ rejected in each iteration. This was the simplest transformation that produced a significant and satisfactory correction. In practice the transformation is calculated by IRAF's `geomap` and performed by `geotran` routines, because they are simple to integrate with routines used to find the arc lines and create the sub-apertures.

The transformation is calculated for each aperture. The optics are stable enough that the transformation need only to be calculated once per night and applied to all images. While a single transformation could, in principle, be used for repetitive nights, we calculate a transformation for each night as we reduce each night independently. Both arc and science images are corrected following the same transformation, which produces a significant improvement in extracted resolution of ∼15 per cent in the detector corners (see Fig. 4).

4.5 Scattered light

Scattered light is an anomalous, large-scale excess signal originating from the scattering of light in the optical components, the 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 corrected.

As described earlier, fibres are organized into slitlets, with a significant gap between them. These gaps allow us to measure the amount of scattered light in individual images (see Fig. 6). There are also eight 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 to the spectrograph. No signal should be present in such gaps, so all the signal we detect there can be attributed to scattered light and fibre cross-talk. While there are also gaps due to dead fibres, these are narrow and there is no guarantee that the fibre has exactly zero throughput.

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. A normal gap between slitlets is 22.5 pixels wide, measured as the distance between trace centres on each side of the gap. Taking the width of the apertures and the 0.7 pixel buffer

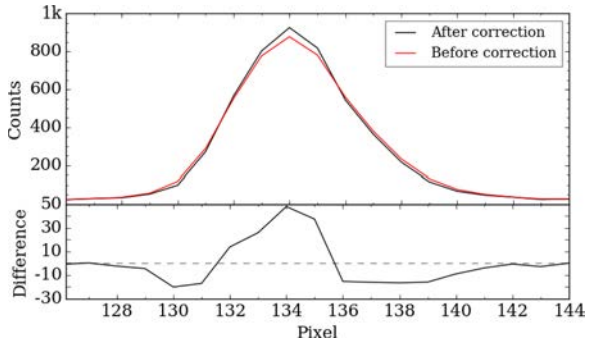

Figure 4. Profiles of the arc line in the top-left part of the red image before (red) and after (black) the geometric correction is applied (top panel). This arc line is also displayed in the left two panels of Fig. 3. The difference between the two profiles is plotted in the bottom panel. The spectrum has already been collapsed into one dimension in this plot.

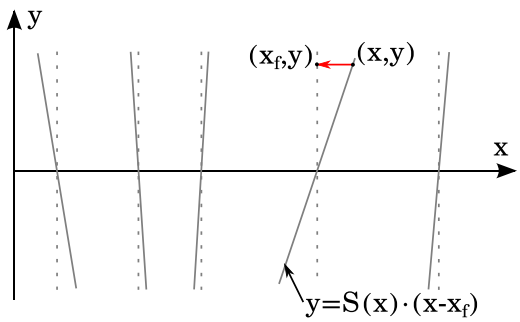

Figure 5. An illustration of the problem, where a varying shear has to be fitted. The x coordinate runs along the wavelength axis, and the y-axis along the sub-apertures (with the middle sub-aperture having the value $y = 0$). Tilted grey lines represent the shear measured from arc lines at different wavelengths, and dashed lines show what the image should look like after the geometric transformation is applied. The equation of each solid grey line is $y = \mathrm{S}(x) \cdot (x - x_f)$, where the shear $\mathrm{S}(x)$ varies from line to line.

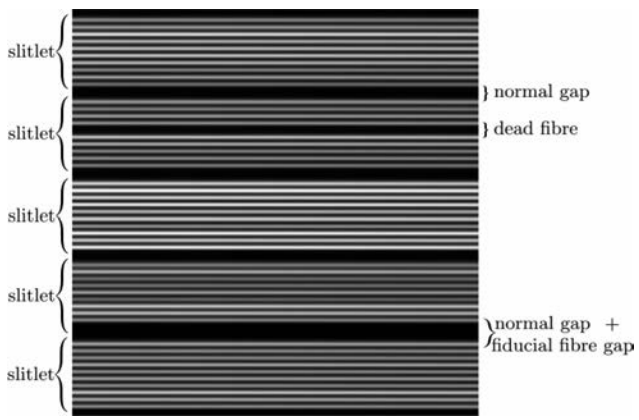

Figure 6. Slitlets and gaps between slitlets presented in a flat-field image. Size of this image excerpt is 500 × 500 pixels.

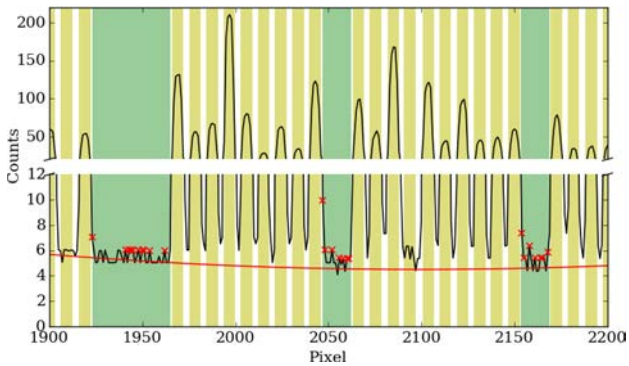

Figure 7. Fitted scattered light. The black line is the median of a spectrum in 70 columns. Yellow regions are extraction apertures, and green regions are gaps. The fitted level of scattered light is the red line. Crosses show points in the gaps that were rejected by the sigma clipping algorithm. Only a small portion of the whole image is shown here.

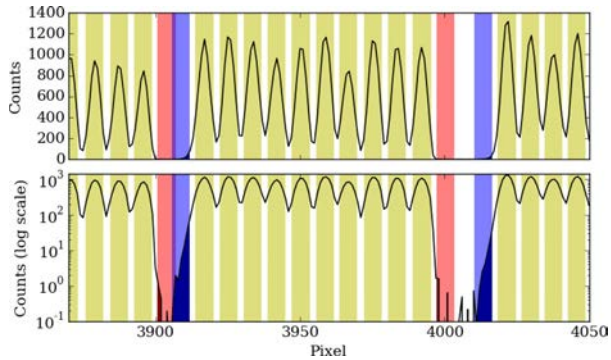

Figure 8. Cross-section of a flat-field showing positions of the phantom apertures in the process of the cross-talk measurement. Yellow regions are regular apertures, red are succeeding phantom apertures and blue are preceding phantom apertures, both placed in a gap. The flux that counts as a cross-talk 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. We show a corner of the flat-field, where the fibre cross-talk is greatest.

into the account, the usable part of the gap is 15.1 pixels wide or more, if there is a fiducial fibre in the gap. As it will become evident in Section 4.6, a large part of any gap must be free of any signal coming from the wings of the spectral traces, so it is only affected by the scattered light.

We use the IRAF's `apscatter` routine to measure, fit, and subtract the scattered light. First, the cross-sections of gaps perpendicular to the dispersion axis are constructed and the signal in 70 consecutive columns is averaged. This averaging produces sufficient 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 4σ and high-rejection limit is 1.1σ). 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 sixth degree Chebyshev polynomial is used to represent the scattered light in both directions and is then subtracted from the entire image.

Fig. 7 shows one of the cross-sections with a fitted polynomial and rejected points. 59 such cross-sections are made over the whole image.

4.6 Fibre cross-talk

Fibre cross-talk[2] is the fraction of the light originating in one fibre and being detected in the region of the CCD (in the aperture) where a neighbouring fibre contributes its light.

Fibre cross-talk is measured by using the slitlet gaps again. Because the fibre cross-talk varies from fibre to fibre, it is impossible to measure it for each fibre individually. However, the variations are small and predictable, so the fibre cross-talk can be measured for each slitlet and interpolated to individual fibres. The amount of fibre cross-talk is independent of the spectrum in each fibre and can be measured in a flat-field image. Two fibre cross-talks must be measured separately: one for the leak from the preceding and one for the leak from succeeding fibres. If a fibre is next to a gap, we assumed there is no cross-talk from that side.

[2] We use the term fibre cross-section for what should be technically called a cross-talk between apertures. Since the effect on the final one-dimensional spectrum is the same for both kinds of cross-sections, we prefer to use a more common term fibre cross-section.

For each slitlet we add two so-called phantom apertures, one before the slitlet and one after the slitlet (as illustrated in Fig. 8), and extract 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. Each slitlet has a preceding and a succeeding phantom aperture, and we measured both cross-talk values. There are 39 gaps in every image, however, some are rejected from this calibration because they include ghosts, as illustrated in Figs 9 and 10. The amount of cross-talk is interpolated with a two-dimensional sixth-order polynomial. This value is then used to subtract the fibre cross-talk from every spectrum. The actual spectrum from the neighbouring fibre to be subtracted is smoothed beforehand with a 5 pixel median box filter to prevent any residual cosmic rays or vertical streaks from propagating into neighbouring spectra.

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

$$S_{\mathrm{F}}^{\mathrm{corr}}(x) = S_{\mathrm{F}}(x) - \left[S_{\mathrm{F}-1}(x)M_{\mathrm{p}}(x,F)\right]\delta(F,F-1) - \left[S_{\mathrm{F}+1}(x)M_{\mathrm{s}}(x,F)\right]\delta(F,F+1),$$

where

$$\delta(F_1,F_2) = \begin{cases} 0, & \text{if there is a gap between } \mathrm{F}_1 \text{ and } \mathrm{F}_2 \\ 1, & \text{otherwise,} \end{cases} \quad (4)$$

where S_{F} is the 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 Fig. 11.

The amount of fibre cross-talk depends strongly on the aperture width, which we have chosen to fix at 6 pixels. This means that some of the signal that does not fall into the aperture is lost, but at the same time the fibre cross-talk is kept low. An aperture width one pixel larger, for example, would mean that the fibre cross-talk is of the same order of magnitude as the signal gained by a wider aperture.

Fibre cross-talk needs only be subtracted once. Further iterations are not needed because only stars in a two magnitudes range are observed at the same time. Any second-order fibre-cross talk is therefore undetectable.

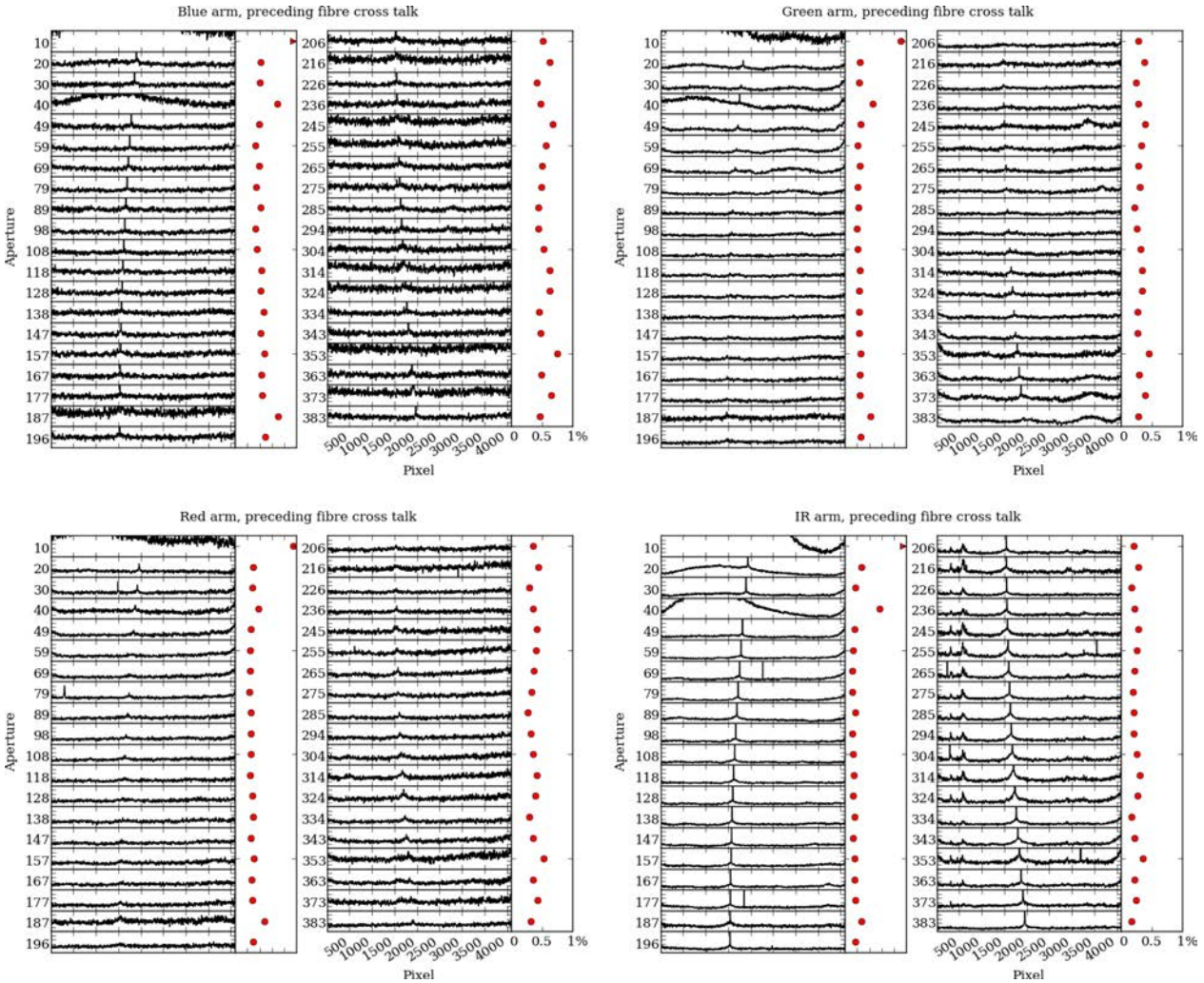

Figure 9. Variation of the preceding fibre cross-talk as it depends on the wavelength (here in pixel scale) and fibre (here given as the aperture number). There are 39 gaps in which we can measure the fibre cross-talk, hence 39 panels. Aperture number is the number of the aperture next to the gap in which the fibre cross-talk was actually measured. Every panel has a vertical scale between 0 and 1 per cent amount of fibre cross-talk. The adjacent panel with red dots shows the average fibre cross-talk for each panel on the left from the dot. 39 panels are split into two columns. A spike in the middle of each panel is a ghost from the 0 order spectrum. It is removed from science frames, but not from the flat-field used here.

The fibre cross-talk measurements demonstrate that the cross-talk changes smoothly from gap to gap, so we can safely assume that it is a result of spectrograph optics and should also change smoothly from fibre to fibre. This means that measuring the fibre cross-talk in the gaps alone is sufficient.

4.7 Wavelength calibration

The wavelength scale is calibrated using spectra from a ThXe arc lamp. One arc spectrum is taken before or after each set of science exposures. It is used as a calibration only for the science exposures observed with that fibre configuration.

4.7.1 The ThXe lamp and line identification

The dominant lines in our lamp belong to Xe, which is not widely used in visual- or near-infrared spectroscopy at wavelengths less than 8000 Å. All the literature line lists (Humphreys & Meggers 1933; Meggers & Humphreys 1934; Humphreys 1939; Zalubas & Corliss 1974; Zalubas 1976; Hansen & Persson 1987; Palmer & R. Engleman 1993; Ahmed, Baig & Suleman 1998; Lovis & Pepe 2007; Redman, Nave & Sansonetti 2014) lacked the desired precision and were found to often be inconsistent with each other. We also see lines in our arc spectra that are not present in the literature line lists. As Fig. A1 shows, we need to use as many of the available lines as possible for calibration. Th lines are generally much weaker than Xe lines, but do have well-known wavelengths. However, there are not enough Th lines to ensure a good wavelength calibration without using the Xe lines as well. The most problematic regions are in the green, red and IR arms where Th lines are either very weak, or totally absent.

To assemble a reliable line list, 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 any concerns about stability and time variation during the exposure. 360 s exposures were made for blue, green, and red arms, which is double the time we use for regular arc exposures. In the IR arm, we took a set of 120 s long exposures, otherwise some Ar lines would saturate and could potentially damage the CCD.

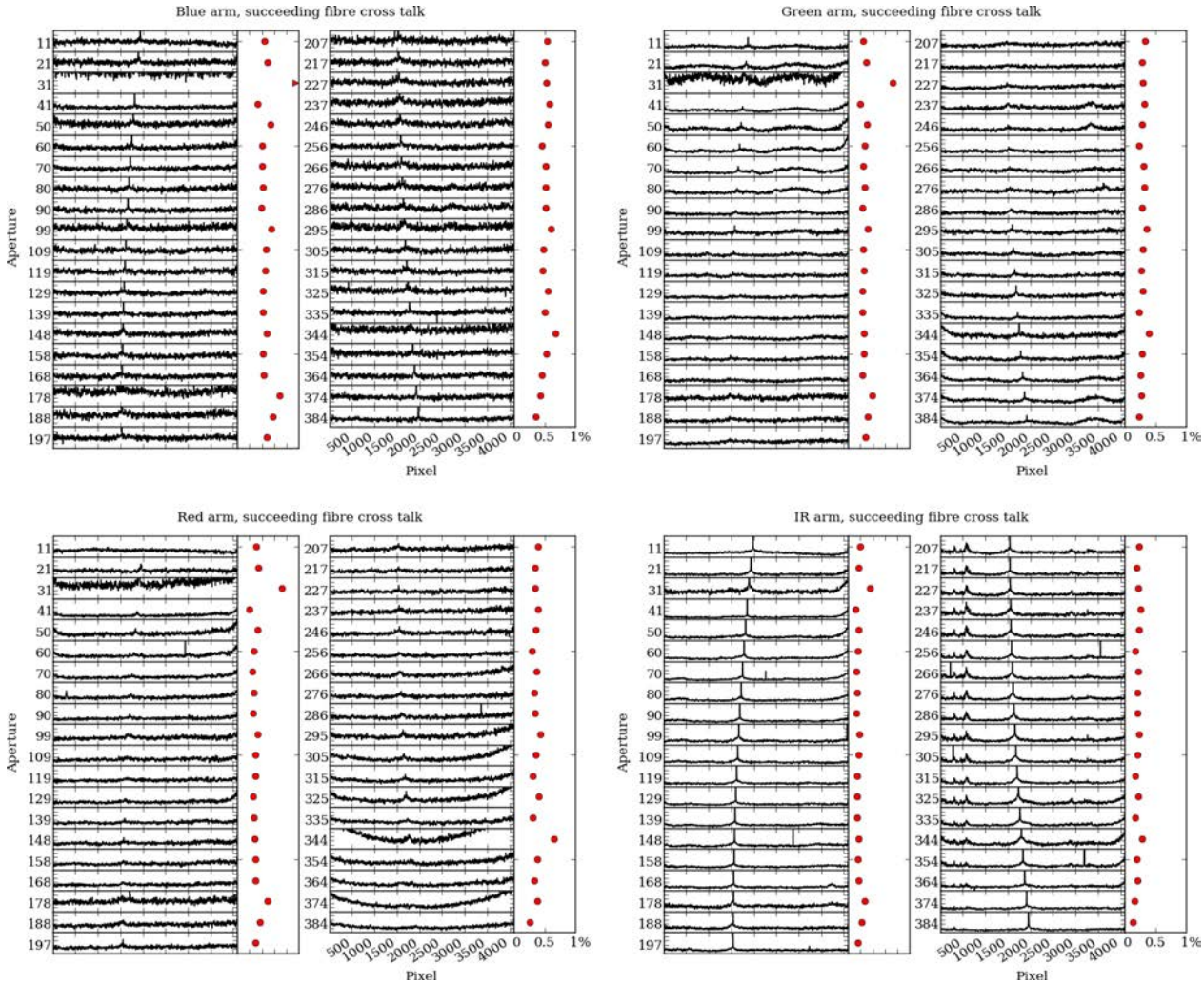

Figure 10. Same as Fig. 9 but for the succeeding fibre cross-talk.

The ThAr+ThXe arcs were wavelength calibrated against the ThAr line list provided in IRAF's `linelist/thar.dat` file (wavelengths for this line list are taken from Palmer & R. Engleman 1993). To assemble the ThXe line list, the wavelengths of Th and Xe lines were measured in the calibrated spectra. We inspected each line by eye and excluded lines that featured peculiar profiles such as 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 Th and Ar lines have a root mean square (rms) residual of 0.0099, 0.0054, 0.0079, and 0.0129 Å in the blue, green, red, and IR arms (respectively).

This two-step process might introduce small systematic errors in the final-measured wavelengths. Fig. 12 shows the discrepancies between our line lists and literature values. The large scatter is primarily due to imprecisions in the line lists from the literature, where mostly Xe lines have wavelengths measured from 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 among themselves, so having a similar shift against our line list is expected. We further demonstrate that the shift is coincidental in Section 6 and Fig. 18. Our final line lists are collected in Appendix.

4.7.2 Wavelength solution

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

To ensure consistent line identification, every arc spectrum is processed with a range of different thresholds for the background taken into the account in the line identification process. With different thresholds different numbers of lines are found. To find the lines IRAF uses our line list and `center1d` algorithm to calculate the central wavelengths of arc lines. The threshold with which the expected number of lines are found is then adopted as a final solution. Detected lines are matched against our line list using our preliminary wavelength calibration (as explained in Section 4.1). A polynomial of the same shape as in the preliminary calibration is then fitted to the detected lines. Lines that are more than 2.5σ from the fit are rejected. Five such iterations are made and the wavelength calibration is then performed on the best-matching lines. We limit the number of rejected lines to 15 per cent of the whole line list, however this limit is rarely reached.

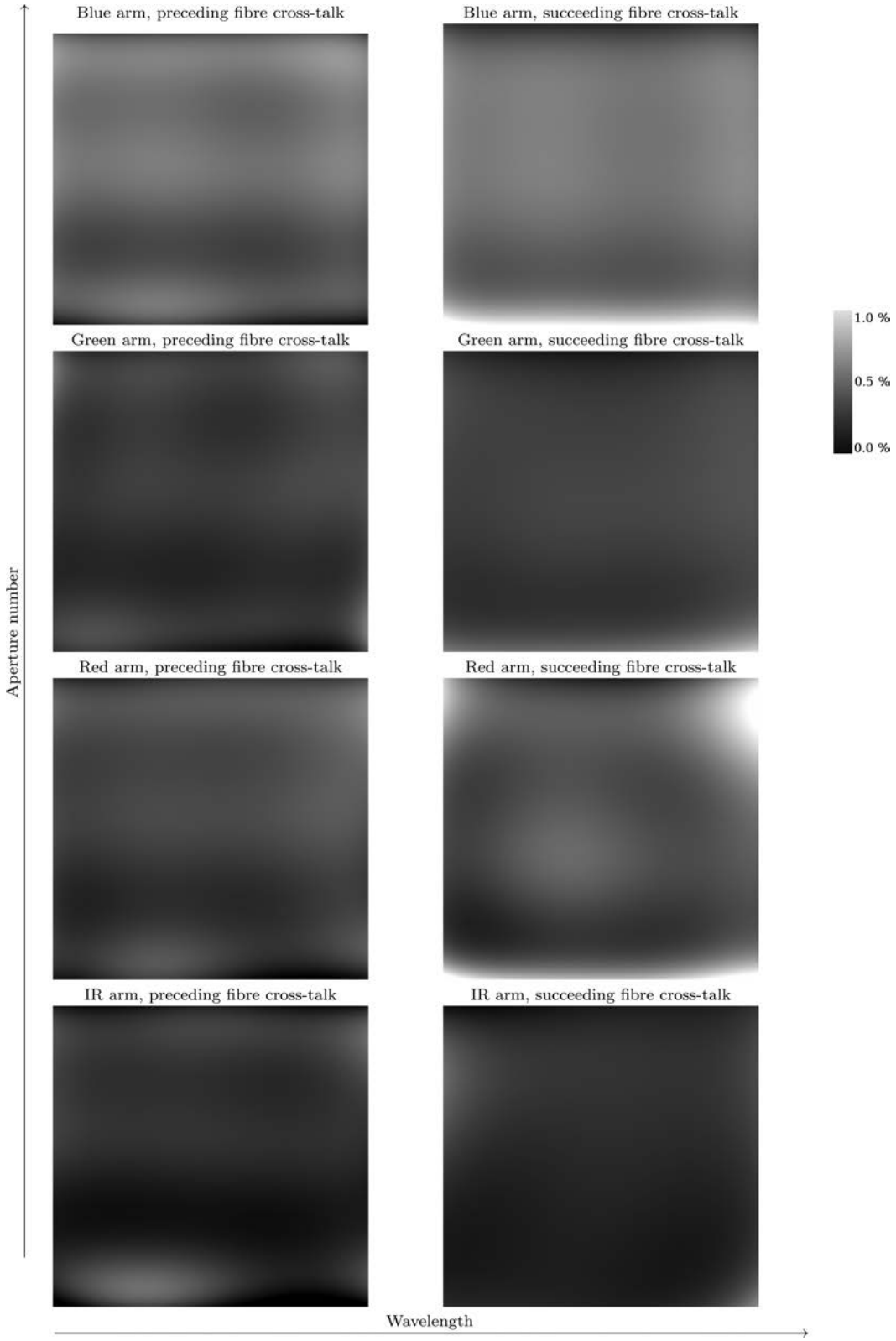

Figure 11. Fitted and smoothed fibre cross-talk for every pixel and for every fibre. There are 4094 pixels per fibre and 392 apertures (eight fiducial fibres are not shown). The grey-scale shows the amount of fibre cross-talk and is normalized to 0 to 1 per cent range.

The fitted wavelength calibration is monitored periodically and if seen to be systematically deviating from the preliminary calibration, the spectrum is tagged, and is then skipped by the radial velocity and stellar atmospheric parameters estimator. This happens in ∼0.9 per cent of the cases for a variety of reasons: low throughput fibre, unpredicted background noise in arc spectra, problems with the arc lamp, and failed solutions due to an arc line remaining undetected in a critical part of the spectrum close to

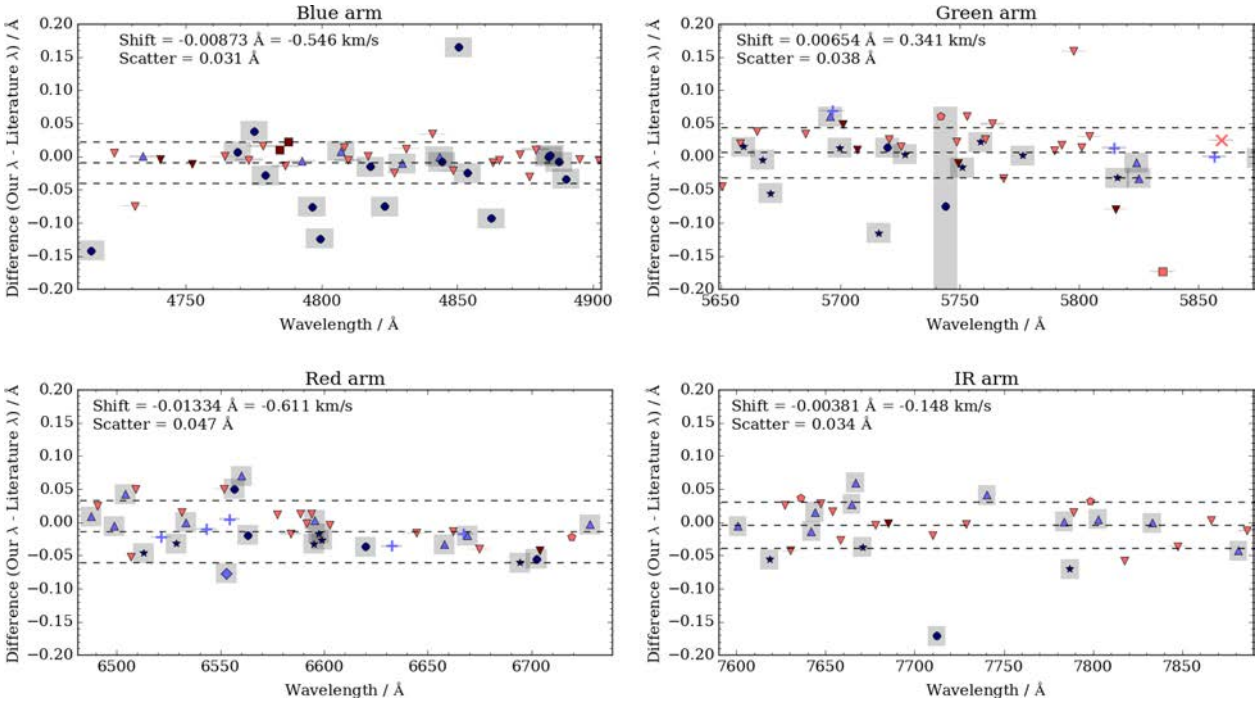

Figure 12. Difference between the wavelengths in our line lists and wavelengths of the same lines in the literature. Dark red symbols represent Th II lines, bright red Th I, dark blue Xe II and light blue Xe I lines. Shape of the symbol marks the literature source (●: Hansen & Persson (1987), ▼: Palmer & 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 A1). Grey 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 A1 are excluded from this calculation.

the edge of the range where the solution cannot rely on any other line.

4.8 Sky subtraction

Out of 392 fibres available in each field, 25 are allocated as sky fibres. These fibres can be different in each field and are positioned as uniformly as possible over the field. Known stars are avoided, but galaxies or other objects are not, so it is possible for a fibre to be positioned at an inconvenient location. Thus, not all 25 fibres can be trusted. Additionally, sometimes not all 25 fibres are allocated on a field due to configuration limitations.

Our sky-subtraction process first models the sky spectrum across the whole field and then subtracts it from all the spectra, as we now discuss.

4.8.1 The sky spectrum

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

When the moon is up, the majority of the sky background in the sky spectrum comes from moonlight. Fields closer than 30° from the moon are avoided, so that moonlight contributes at most 120 counts per hour per pixel to the sky spectrum.

In addition, we observed an unexplained flux of about 10 to 20 counts h^{-1} $pixel^{-1}$ (in blue and IR arm, respectively). The spectrum of this contribution appears smooth. Despite our best efforts, we were unable to identify the source. It does not show any telluric absorption lines, so it must originate from somewhere near the telescope. This unexplained flux is best seen in the IR panel of Fig. 13, where saturated telluric absorption lines do not reach zero flux.

In a few fields close to the Galactic plane we observe emission lines coming from hot gas in the Milky Way itself. In these cases the sky subtraction cannot be trusted at wavelengths near these lines, as the nebulae of the Milky Way are much more structured than can be sampled by 25 sky fibres. Fortunately, the GALAH survey typically observes regions far away from the Galactic plane ($|b| > 10°$, see Martell et al., in preparation), and so such occurrences are very rare. Automated flagging of this condition has not been implemented in the pipeline, however the pipeline does provide a spectrum before the sky subtraction, so the user can decide whether or not the subtraction is still usable for their particular applications.

4.8.2 Subtraction

We need to be able to interpolate a sky spectrum over the whole 2° diameter field using 25 sky spectra (some of which may be contaminated by background sources). We must also account for the changing resolution delivered as a function of wavelength and position on the CCDs (see Section 7). The latter is largely solved by correcting the optical aberrations. Only a small variation of roughly 3 per cent remains (see Figs 21 and 22). This is small enough (compared to the intrinsic relative noise of the sky emission lines) that taking it into account is not necessary. Resolution changes more significantly with wavelength, but this change is the same for both

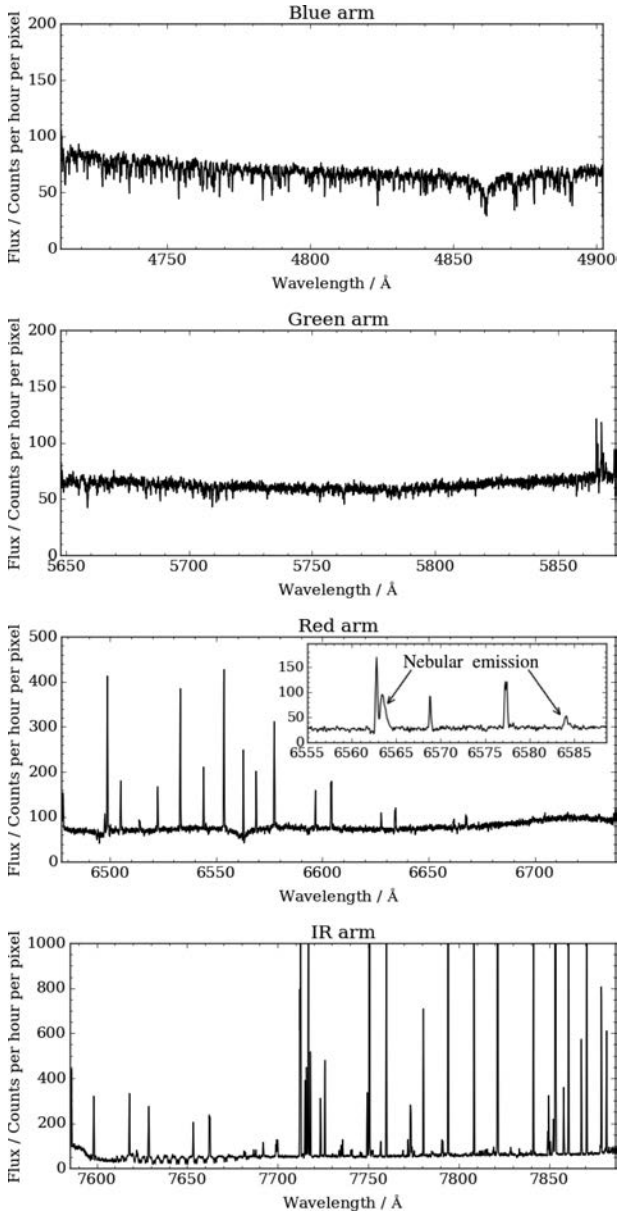

Figure 13. Spectra of a moon-lit sky on 2015 June 6. A typical spectrum of the night sky is shown for all four arms. The main contribution to the sky spectrum in this case is the moon-light. Almost all of the emission lines come from the air glow. The inset is from a different night than the other spectra, when there was no moon above the horizon and in 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 that appears smooth and for which we could not find a good explanation.

sky fibres and object fibres so the subtracted sky spectra always have the correct resolution at every wavelength.

First, we measure the fibre throughputs from a flat-field. Excluding dead fibres, the throughputs of fibres can vary up to 50 per cent for different fibres (Simpson et al. 2016). As noted above, the flat lamps do not illuminate fibres evenly; the illumination depends on the position and orientation of each fibre. However, a flat-field is taken for every observed fibre configuration, so combining flat-fields from different configurations is enough to average out the variations in the fibre illumination. Fibre throughputs are therefore measured

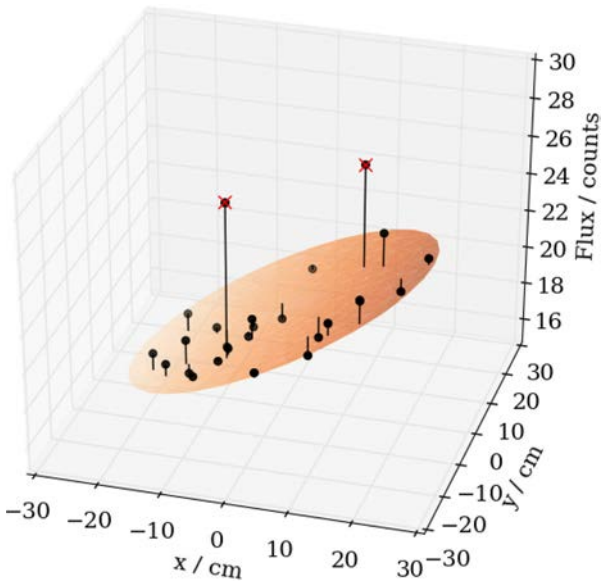

Figure 14. Fit of a model sky plane to sky fibre fluxes. Measured fluxes are represented by black points. After the fitting, two points were rejected and are crossed out in the image. The plane shows the size of the plate where fibres can be positioned. It is colour coded by the mapped flux ranging from 16.5 to 23.8 counts pixel^{-1}. The x and y axes show position on the actual plate. This example shows red arm data.

from an averaged flat-field and throughput is parametrized as a function of wavelength using fifth-order Chebyshev polynomials. The fibre with the highest throughput has its throughput set to one and all the spectra are normalized according to this scale with their corresponding polynomials. Absolute values of the throughputs are not relevant because we do not aim to produce flux-calibrated spectra.

Next, all 25 sky spectra are averaged with a rejection of values that are more than 3σ away from the average into a single spectrum (five iterations are made). Spectra from ill-positioned fibres or any remaining cosmic rays, hot pixels, and vertical streaks are therefore removed from the averaged spectrum. This results in a high-SNR sky spectrum, so the subtraction will introduce less error.

We then map the sky flux 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: the 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 fit are performed, where points more than 3σ away from the plane are rejected in each step. This again removes outliers due to ill-positioned fibres. An example of such plane is shown in Fig. 14.

We experimented with more complicated functions, like warped plane, where the tilt of the plane changes along one direction. This proved to be too complicated to fit automatically to 25 points while allowing the rejection of outlying points.

The final step is then to take the position of each object fibre, calculate the model sky flux from the value of the model plane at that position, scale the high-SNR sky spectrum accordingly and subtract it from the spectrum for that fibre. Fig. 15 shows an example of a sky-subtracted spectrum in the IR arm where sky emission lines are strongest.

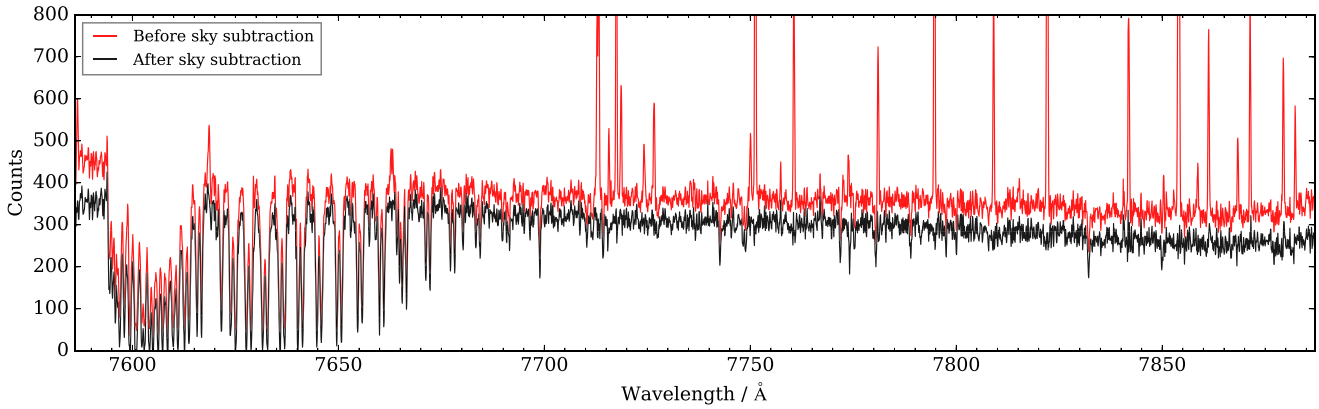

Figure 15. IR arm spectra from 2014 September 9 show the spectrum before the sky subtraction (red) and after the sky subtraction (black). The spectrum has a low SNR (SNR = 25) so the sky emission lines are clearly visible. Note that also some continuum sky-light was subtracted.

4.9 Telluric absorption line removal

Removal of telluric absorption lines usually requires additional observations of standard stars, which is time-consuming and not feasible for GALAH. It would also require us to model the absorption line profiles across the sky and with time to properly account for the effect of telluric absorption on GALAH science spectra. However, the GALAH wavelength ranges are impacted by telluric absorptions and do have to be corrected. We correct GALAH science spectra by fitting the telluric lines directly in our observed spectra.

While telluric absorption lines are present in all four arms, they are only strong enough to be detectable and corrected in the red and IR arms. The blue end of the IR arm is in large part truncated by the *A* band, a series of O_2 vibrational lines. Many of 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 so that the astrophysically interesting oxygen triplet lines lie in the centre of the range, which made the inclusion of a large part of the *A* band. There are no features essential to the GALAH survey 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 absorption is fitted with European Southern Observatory's MOLECFIT software (Kausch et al. 2015; Smette et al. 2015), release 1.1.1. MOLECFIT fits synthetic telluric absorption lines to spectra. To fit the telluric absorption spectrum to each spectrum takes around a minute on a top-end desktop computer, so it is infeasible 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 for an entire field. Our fields are observed at airmass low enough that no telluric line (with the exception of the *A* band) will become optically thick in any part of the field. For each fibre, we then calculate the difference between the average airmass and its actual airmass, and scale the fitted telluric spectrum accordingly. The individual spectra are then divided by scaled telluric spectra. An example of this process is shown in Fig. 16.

4.10 Barycentric velocity correction

The wavelength calibration is adjusted for the barycentric velocity correction. The correction takes the effect of the Moon's motion and rotation of the Earth into 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

Our data start with an uncertainty determined by the photon-counting uncertainties in every pixel, and almost every step described above introduces further uncertainty and/or scales the spectrum away from a photon-counting distribution. We therefore propagate an initial photon-counting (i.e. Poisson) standard deviation for each pixel through the analysis to produce a so-called error spectrum, or uncertainty spectrum that describes the standard deviation of each pixel in the final spectrum.

5 ORGANIZATION OF REDUCED DATA

5.1 Individual spectra

Every reduced spectrum is named with a 16 digit number:

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

(a) Observation date in the format two digits for the year, two for the month, and two for the day. This is the date when the night starts and 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 three exposures).

(d) Three digit pivot number.

(e) Arm number. 1 is blue, 2 is green, 3 is red, and 4 is the IR arm.

This name is unique and is used throughout, rather than star names or input catalogue numbers. Additionally, due to the exceptions like pilot programmes and programmes that do not use the GALAH selection function, but share the same observing routine, not all targets have input catalogue numbers.

There is one FITS file created per reduced spectrum. It contains four extensions: (i) the reduced spectrum itself, without any normalization; (ii) the error spectrum, which gives the relative uncertainty of the reduced spectrum. (iii) the spectrum without the sky subtraction 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); (iv) the relative error spectrum giving the uncertainty of the spectrum without the sky subtraction and with no telluric absorptions fitted.

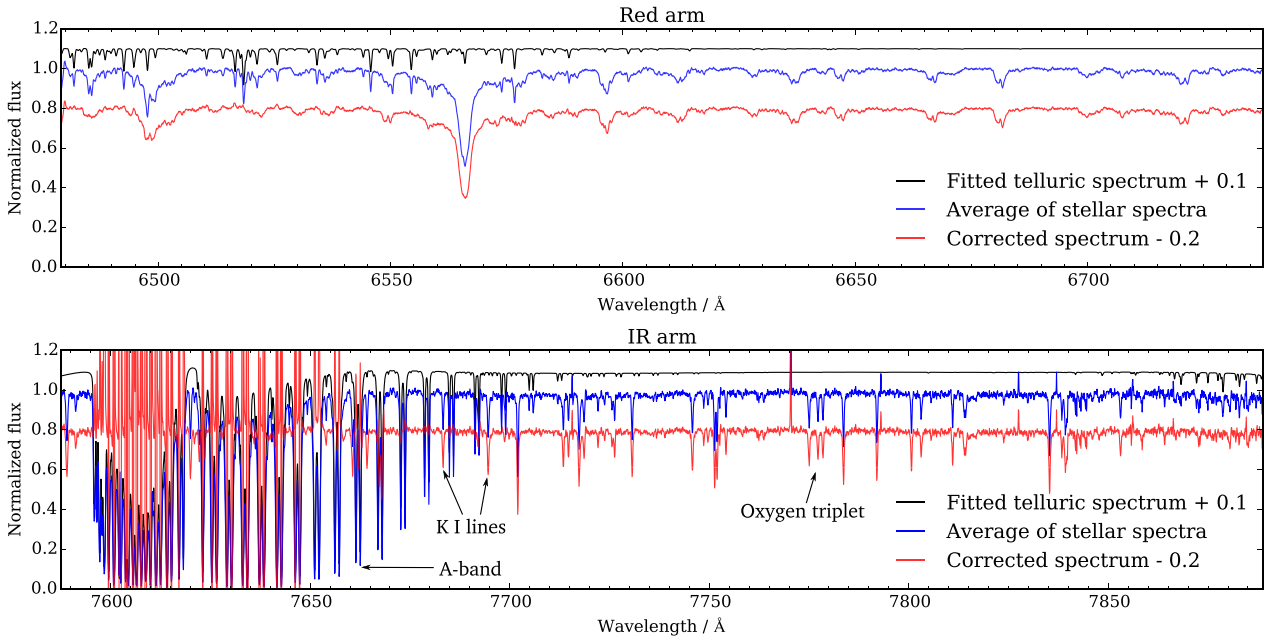

Figure 16. Fitting and removal of telluric absorption lines. Plots show the fitted and corrected spectra in the red arm (top panel) and in the IR arm (bottom panel). The average spectrum of all objects in one image is plotted in blue. Black spectrum is the fitted telluric absorption spectrum. The corrected 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. Features discussed in the text (oxygen triplet, K I lines and the *A* band) are marked.

Until the final step of the reduction, each spectrum has a wavelength calibration saved as a polynomial giving a relation between pixels and wavelength. This is impractical in the next steps – combining individual spectra and then estimating radial velocity and stellar atmospheric parameters, so the wavelength calibration is linearized by linearly interpolating the values between pixels. Median dispersion is used for the linearized wavelength calibration, so the number of pixels per spectrum remains roughly the same after this process. The wavelength calibration is the same for all four extensions. It is linearized and in a standard FITS format given by 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*

While usually three consecutive 20-min exposures are made for each field, this is not always the case as there can be more, or less, depending on weather conditions (as decided by the observer at the telescope). A flat-field and arc frame are also made with the same fibre setup as the science exposures for each field. This 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 makes one observation used for further scientific analysis. Individual uncombined spectra are retained, but usually not used for science purposes. The cob_id is simply the combination of the six-digit date marker and the run number of the first exposure made in this block, regardless of 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 they will be treated 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 are treated separately.

A new name is assigned to combined spectra:

$$\underbrace{1402310042}_{a}\underbrace{01}_{c}\underbrace{125}_{d}\underbrace{2}_{e}.\text{fits}.$$

Instead of the date and run marker, we here use the cob_id (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, we use two combining methods, which are reflected in the file name as follows.

(1) All the science exposures of the same object with the same cob_id are summed together. Error spectra are combined appropriately, so they still give relative uncertainties.

(2) All the science exposures with the same cob_id are combined as in the case 1 above. Additionally, all the exposures of the same object made at earlier times are also added, even though they have different cob_ids. Error spectra are combined appropriately, so they still give relative uncertainties. We emphasize here that this combines an exposure of an object with all previous exposures. The cob_id of this output spectrum is the cob_id of the last combined exposure. For example, if an object is observed on 3 d, there are two resulting spectra with combining methods set to 2. The first combines the exposures from the first and second day and carries the cob_id of the second day. The second output spectra combines the exposures from all three days and has the cob_id of the third day. To get the version of the spectrum that combines all previous exposures, select that with the latest cob_id.

Sometimes another number called sobject_id is used. It is the same as the name of the spectrum described above, but drops the last digit (denoting the arm), so it stands for the object as a whole and not for individual spectra for each arm. An sobject_id for the above example will therefore be:

$$\underbrace{1402310042}_{a}\underbrace{01}_{c}\underbrace{125}_{d}$$

where *a*, *c*, and *d* mark the same parts of the name as before.

6 ESTIMATION OF RADIAL VELOCITIES AND STELLAR ATMOSPHERIC PARAMETERS

After the reduction of spectra we run a code that provides first estimates of radial velocity and three basic atmospheric parameters for each star. The spectra are delivered to the collaboration only after this step is performed, so we present this software code together with the reduction pipeline. Radial velocities and stellar parameters are calculated using the blue, green, and red arms only. The fourth arm is omitted as it contains significant amount of telluric contamination and comparatively less parameter-sensitive information than the other three arms.

Radial velocities are calculated by cross-correlating the observed spectrum with a set of 15 AMBRE model spectra (De Laverny et al. 2012). 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 and 0 dex, respectively.

Normalization is done by fitting a univariate spline function over the entire observed spectrum. 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. This is done to reduce the noise in the cross-correlation peak. The normalized spectrum is then cross-correlated with the 15 models to determine the radial velocity from this particular arm. Cross-correlation function is calculated in steps of one pixel, so the exact radial velocity is calculated by fitting the highest peak in the cross-correlation function. The cross-correlation peak is fitted with a second-degree polynomial and we adopt the maximum of the fit as the cross-correlation coefficient. Only model spectra which have cross-correlation coefficient greater than 0.3 (the coefficient ranges from 0 to 1, with 1 being the perfect correlation) are included in the radial velocity determination.

The radial velocity for each arm is a combination of the acceptable model spectra, weighted by their cross-correlation coefficient. The radial velocities from all three arms are averaged to give the final radial velocity, and the standard deviation between the arms is adopted as the uncertainty. The uncertainty is not derived from the peak width of the cross-correlation function, because of large dispersion and cannot be calculated accurately from a sample of only 15 cross-correlation functions. If one arm has a radial velocity lying more than two times further than the difference between the other two, it is excluded from the final radial velocity. Fig. 17 shows the comparison between radial velocities from this work and from literature for four well-studied clusters. With the exception of M67, our radial velocities for each cluster agree with literature values within their cited uncertainties. For M67, the peak of our velocities agree well with the literature value, however, the observed dispersion is large compared to literature values because it includes possible non-members, an unknown fraction of spectroscopic binaries, and has not been corrected for the individual velocity errors, which for this cluster, are comparable with the intrinsic cluster velocity dispersion.

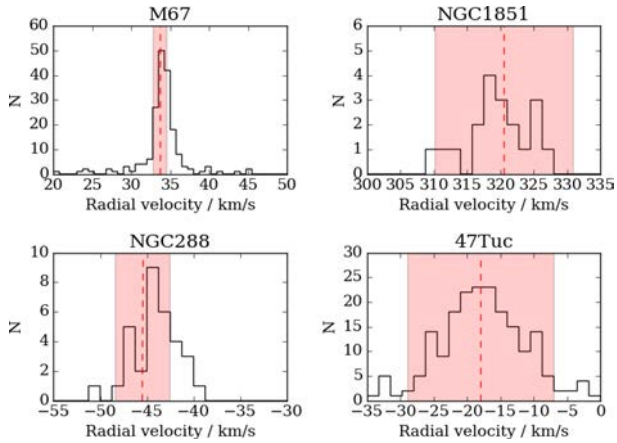

Figure 17. Radial velocity distributions of four well-studied clusters. The red dashed lines are literature values, the red boxes indicate their associated dispersions. For each cluster, we first give the mean radial velocity and dispersion from the literature and then our own values. Top left panel: M67, 195 stars. Literature values: 33.64 and 0.854 km s^{-1} (Geller, Latham & Mathieu 2015), this work: 33.71 and 8.02 km s^{-1}. Top right panel: NGC1851, 19 stars. Literature values: 320.5 and 10.4 km s^{-1} (Harris 1996), this work: 319.52 and 4.75 km s^{-1}. Bottom left panel: NGC288, 32 stars. Literature values: −45.4 and 2.9 km s^{-1} (Harris 1996), this work: −44.22 and 2.28 km s^{-1}. Bottom right panel: 47 Tuc, 196 stars. Literature values: −18 and 11 km s^{-1} (Harris 1996), this work: −17.42 and 6.61 km s^{-1}. It is evident that our radial velocities match nicely the literature radial velocities for these four clusters. We attribute the discrepancies in the radial velocity dispersions to contamination by field stars (in M67), small number of observed stars (in NGC1851), and cluster centres not being included in GALAH observations (NGC1851 and 47Tuc).

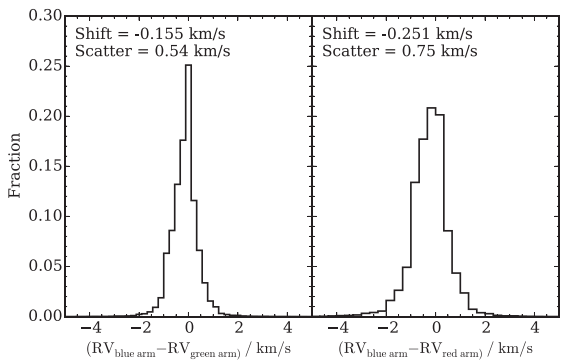

Figure 18. Difference between the blue arm and green arm radial velocity (left-hand panel) and the difference between the blue and red arm radial velocity (right-hand panel). Radial velocity is not measured from the IR arm. Shift from zero difference and scatter of the distribution are given in each panel. Radial velocities of ∼124 000 stars were used to produce this plot.

Fig. 18 shows the statistics for the differences between the radial velocities measured in each arm. There is a systematic shift between each arm. It does not correspond to the shift we see in Fig. 12, concluding that the shifts shown in Fig. 12 are a consequence of imprecise literature values for a large number of lines. We attribute the non-zero radial velocity differences in Fig. 18 to systematic errors in the radial velocity calculation scheme and the initial ThXe line list we composed, of which the issues are propagated to every radial velocity calculated. However, the shift is almost two orders of

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

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 3. Parameter summary of the model grid.

Parameter	Range	Increments
$T_{\rm eff}$ (K)	2500–8000	200 ($T_{\rm eff} < 4000$)
		250 ($T_{\rm eff} > 4000$)
$\log g$/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])

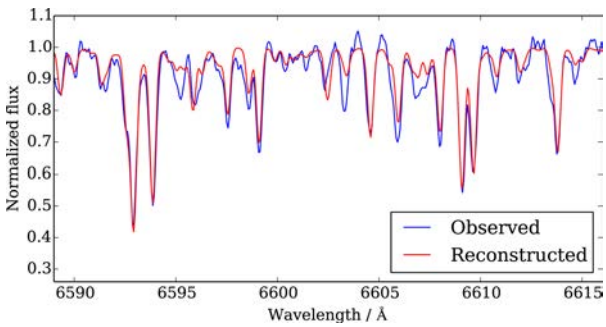

Figure 19. The comparison between the reconstructed (red) and the observed (blue) spectra for a star. Most spectral features are well reproduced, with only a few lines (e.g. line at 6603.5 Å) with erroneous strengths. Such deviations are caused by inaccurate line lists used to produce the synthetic spectra.

magnitude smaller than the resolution of the spectrograph, meaning that our wavelength calibration method works.

Stellar parameters ($T_{\rm eff}$, $\log g$, and [Fe/H]) are derived by globally fitting of the observed spectra with a grid of 16 783 AMBRE spectra (De Laverny et al. 2012), which are convolved down to the average resolution of each CCD. Tables 2 and 3 contain the wavelength and parameter ranges of this grid. The aim of this procedure is to provide initial stellar parameters, which are used as the starting model to help speed up the stellar parameters and abundance pipeline. This global fitting method is both robust and efficient, taking under a second per star to converge.

The radial velocity corrected spectrum is normalized by fitting third- (for blue and green arms) and fourth- (for red and IR arm) order polynomials to predetermined continuum regions. These regions were determined by examining high resolution spectra of the Sun, Arcturus, and μ Leo. Normalized spectra are interpolated on to the common wavelength grid as the model. We select the 10 closest-model spectra in Euclidean space to the observed spectrum (by computing the L^2 norm between the observed spectrum and each model spectrum) and reconstruct the original spectrum using a linear combination of these 10. The stellar atmospheric parameters are the same linear combination of the parameters of the 10 closes-model spectra. 10 spectra proved to be a safe number to use considering the size of the synthetic grid and quality of our data. Fig. 19 shows an example of the reconstructed and the observed spectra.

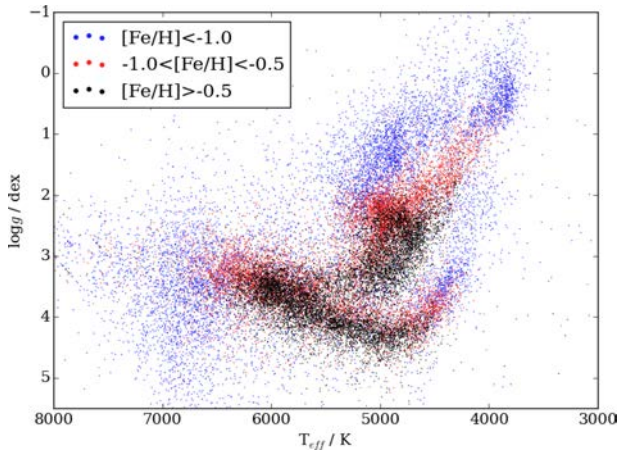

Figure 20. The spectrometric Kiel diagram of 10 000 randomly selected GALAH stars with metallicity binned into three bins.

Fig. 20 shows the Kiel diagram of 10000 GALAH stars with parameters derived from this method. We note the erroneous upturn at the lower part 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 (e.g. Worley et al. 2016).

7 RESOLUTION MAPS

Resolution of reduced spectra is not self-evident, despite the information provided by the instrument designers (Sheinis et al. 2015). The star light passes through many optical elements and computational processes, so it is not trivial to calculate the resolution of final reduced spectra. It has to be measured from actual observed spectra. It is, however, critical that the real resolution is known by analysis methods. We therefore measured the resolution by two methods: from arc spectra and from spectra produced by an optical comb (Steinmetz et al. 2008).

7.1 Resolution maps from arc lines

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 is to use sky emission lines, also unresolved but weaker than arc lines and found only in parts of the red and IR arm.

We measured the 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:

$$R = \frac{\lambda}{\rm FWHM}, \tag{5}$$

which was calculated for each measured line. A sixth 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 Figs 21 and 22.

Due to concerns that the emerged pattern is distorted by blends of unresolved arc lines, we tried to verify the measured resolution from measurements of profiles of sky emission lines. Because sky emission lines are weak in individual spectra, we used combined sky spectra, like the one in Fig. 13. The result is shown in Fig. 22

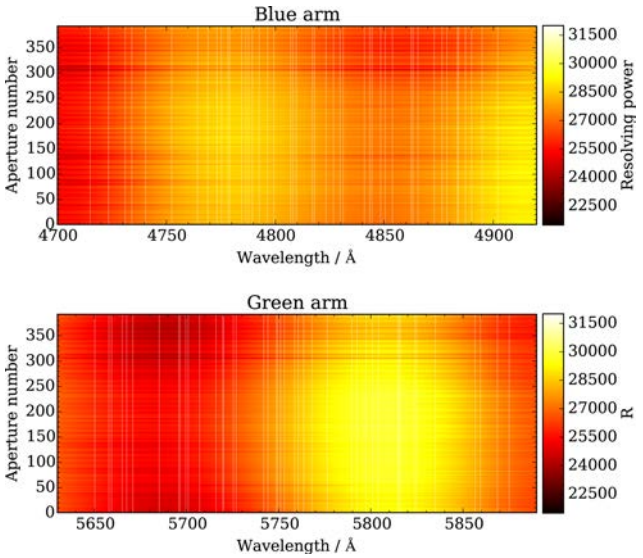

Figure 21. Resolution maps for the blue arm (top) and the green arm (bottom). The maps show the colour-coded resolving power as it depends on the wavelength and aperture number. White vertical lines show arc lines that were used to produce these maps. Map is composed of sixth degree polynomials that represent the resolving power for each aperture. Polynomials were fitted to measured arc lines widths for each aperture independently.

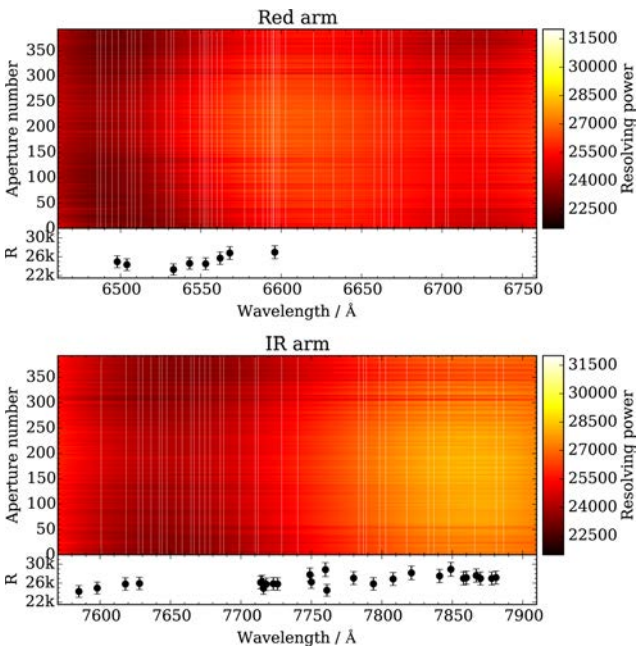

Figure 22. Resolution maps for the red arm (top) and the IR arm (bottom). The maps show the colour-coded resolving power as it depends on the wavelength and aperture number. White vertical lines show arc lines that were used to produce these maps. Map is composed of sixth degree polynomials that represent 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.

under the red and IR maps. There are no usable sky emission lines in the green and blue arms and only a few in the red arm. We observe a correlation in the resolution measured in both ways, so we believe that the measured resolution profiles are the consequence of the instrument and not of inconveniently blended arc lines. We test this further with the help of a photonic comb in the following section.

The measured resolution is in a rough agreement with the resolving power of 28000, for which the instrument was designed. The 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.

7.2 Resolution maps from the photonic comb

In 2016 March and April, we were given 4 d at the AAT to try coupling a photonic comb to the 2dF for the purpose of measuring resolution, check our method of wavelength calibration, and test the capability of a photonic comb to serve as a calibration for observations in the future. We were unable to take data of sufficient quality to challenge the wavelength calibration method presented in Section 4.7, but the data was suitable for the resolution measurements.

The photonic comb spectrum is generated by coupling white light into a single-mode fibre based broad-band Fabry–Perot etalon manufactured to have a finesse (ratio of the etalon's transmission peak separation to the width of the peaks) of 40 and a 453 μm cavity. This produces a peak separation of 4 Å at 7800 Å and 2.5 Å at 5700 Å both with a width of approximately 1.1 Å. The etalon is enclosed with a small thermoelectric cooler that is used to stabilize the transmission spectrum. The small size of the cavity allows for relatively quick changes (multiple degree changes can settle in less than a second). To stabilize the photonic comb spectrum, we use a Doppler-free saturation absorption spectroscopy set-up (Preston 1996) that allows simultaneous measurement of the rubidium D_2 hyperfine transition lines and a transmission line of the photonic comb using a tunable laser. This rubidium transition is known to be extremely stable and is an excellent absolute wavelength reference. The photonic comb spectrum is transmitted to the spectrograph using a series of single-mode fibre switches to swap the input and output of the etalon between a white light source or the tunable laser and to feed either the spectrograph or Rb locking setup, respectively. As a precaution the setup was limited to monitoring the etalon in-between exposures to avoid coupling the tunable laser into the spectrograph. These measurements indicated stability on the order a few m/s over each exposure.

The photonic comb produces Lorentzian profile lines with a FWHM of 0.15 Å in the IR arm, which is less than the resolution. However, the wide Lorentzian wings are resolved. Even more, they are detectable in the whole region between two peaks in a reasonably exposed spectra with more than a few thousand counts in the peaks. When measuring the resolution, the spectra should be deconvolved with the light source spectral profile. Resolving the Lorentzian profiles allows us to skip this step and fit Voigt profiles (a convolution of a Gaussian and Lorentzian profile) instead. The Voigt profile can be conveniently written as

$$V(x; A, x_0, \gamma, \sigma) = A\frac{\mathrm{Re}\,(w\,(z))}{\sigma\sqrt{2\pi}},$$
$$w(z) = \exp\left(-z^2\right)\mathrm{erfc}(-iz),$$
$$z = \frac{(x - x_0) + i\gamma}{\sigma\sqrt{2}}, \tag{6}$$

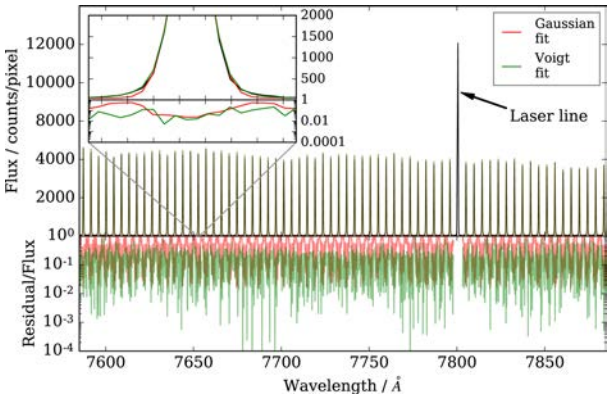

Figure 23. Spectrum produced by the photonic comb. The marked line is the Rb laser line. The laser does not shine directly into the fibre, but due to its strength some light scatter of the dome and into the fibres. The inset shows one typical line in detail. Red and green lines show fitted Gaussian and Voigt profiles. Only the IR arm spectrum is shown here. Green and Red arms spectra look similar, but have somewhat less signal and no laser line of course. Blue arm spectra are useless for our application because the photonic comb produces too weak signal at those wavelengths.

where A is the amplitude, x_0 is the centre of the profile, γ is the parameter of the Lorentzian part, and σ is the parameter of the Gaussian part of the Voigt profile. The resolving power can then simply be calculated by adopting σ:

$$R = \frac{\lambda}{2.3548\,\sigma}, \tag{7}$$

Fig. 23 shows one such fit. Note that the central part of the peak is fitted equally well by a Gaussian and Voigt profile, while wings can only be fitted by a Voigt profile. This means that the wings give all the information about the Lorentzian part of the Voigt profile. The resolution in Fig. 24 is measured as a FWHM of the Gaussian part of the Voigt profile.

Fig. 24 can be compared to Figs 21 and 22. While resolution measured via arc lines and the photonic comb is consistent in green and red arms, the shape is different in the IR arm. The wavy pattern along the wavelength axis that we see in Fig. 22 is absent in Fig. 24. We conclude that the part with lower resolution at around 7650 Å that is making the pattern wavy is caused by the arc lines that are actually unresolved blends of several lines.

8 CONCLUSIONS

In this paper, we presented the current state of the IRAF reduction pipeline for the GALAH survey. The code incorporates all major steps needed to extract a one-dimensional spectrum from the acquired images and correct for known aberrations. Spectra are wavelength calibrated, but not flux calibrated, because this is not part of GALAH's science goals, and the chosen observing strategy therefore does not support it. The targeted precision of the calibrations and corrections that are made also driven by GALAH's scientific goals. The abundance pipelines require the wavelength calibration to be only slightly better than 1 km s^{-}1, which is achieved, but the requirement is not surpassed. Therefore, the GALAH project does not yet serve as a very precise radial velocity survey. The same principles guide the atmospheric parameters and radial velocity estimator – it is designed to only give a first approximation of stellar atmospheric parameters that more advanced analysis pipelines can use.

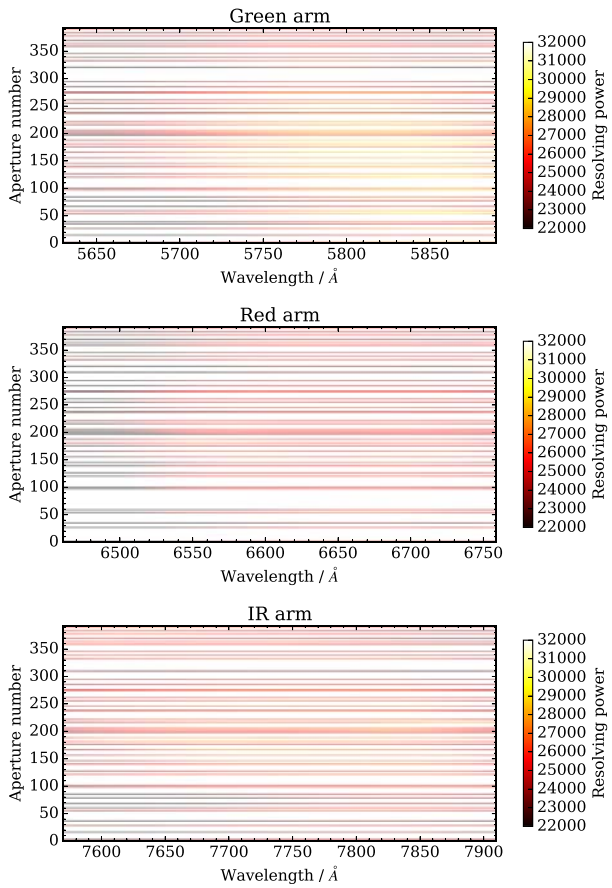

Figure 24. Resolution maps for all three analysed arms with fibres from plate 0. Only 43 fibres were probed. White gaps represent fibres that were not probed.

Because the survey is still ongoing and observations will continue for some time, this will certainly not be the final version of this code. The observing strategy is expected to remain the same, but there is a possibility that new observing projects will be incorporated into the GALAH workflow. The GALAH survey is the first project to use the new HERMES spectrograph at the AAT. We therefore present some performance analysis (resolution maps) and give references (ThXe arc line lists) that may be useful to other users of HERMES. Apart from simulations and laboratory tests, results presented in this paper are the first use of the HERMES spectrograph in a real-life application.

Maintenance of the HERMES spectrograph is planned for the near future. This aims to fix problems with the slightly out-of-focus red arm CCD. The source of the vertical streaks will also be sought and possibly eliminated by replacing some optical elements (Martell et al., in preparation). This will change the performance of the spectrograph, but should not impact the IRAF reduction pipeline too much. While it is expected that some fine tuning may be required after this maintenance, the overall process will remain the same. The resolution will also have to be measured again, as it might change after some optical elements are replaced. We plan to utilize an optical comb to do this job, because it proved to produce much more reliable and controlled spectra than the ThXe or ThAr lamp. In order to keep a consistent quality control, we plan to probe all fibres on both plates before and after the maintenance. We also plan to recalibrate the Xe line wavelengths with the optical comb to improve the wavelength calibration to a level where the precision

of measured radial velocities will be limited by the spectral noise and not the wavelength calibration.

ACKNOWLEDGEMENTS

We thank the anonymous referee for valuable comments and suggestions that improved the quality of the paper and its presentation.

The data in this paper were based on observations obtained at the Australian Astronomical Observatory as part of the GALAH survey (AAO Programmes 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, Astrophysics Source Code Library, record ascl:1505.015
Ahmed M., Baig M. A., Suleman B., 1998, J. Phys. B: At. Mol. Opt. Phys., 31, 4017
Blecha A., Cayatte V., North P., Royer F., Simond G., 2000, in Iye M., Moorwood A. F., eds, Proc. SPIE Conf. Ser. Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. SPIE, Bellingham, p. 467
Colless M. et al., 2003, preprint (astro-ph/0306581)
Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397
Czesla S., 2015, PyAstronomy, available at: https://github.com/sczesla/PyAstronomy
De Laverny P., Recio-Blanco A., Worley C. C., Plez B., 2012, A&A, 544, A126
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Dominquez Palmero L., Jackson R., Molaeinezhad A., Fariña C., Balcells M., Benn C. R., 2014, in Peck A. B., Benn C. R., Seaman R. L., eds, Proc. SPIE Conf. Ser. Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V. SPIE, Bellingham, p. 91492J
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Geller A. M., Latham D. W., Mathieu R. D., 2015, AJ, 150, 97
Hansen J. E., Persson W., 1987, Phys. Scr., 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, available at: https://github.com/wlandsman/IDLAstro
Lovis C., Pepe F., 2007, A&A, 468, 1115
Luo A.-L. et al., 2015, Res. Astron. Astrophys., 15, 1095
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., Engleman R., Jr, 1993, Technical report, Atlas of the Thorium Spectrum. Los Alamos National Lab, Los Alamos, NM, USA
Preston D. W., 1996, Am. J. Phys., 64, 1432
Redman S. L., Nave G., Sansonetti C. J., 2014, ApJS, 211, 4
Sheinis A. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 035002
Simpson J. D. et al., 2016, MNRAS, 459, 1069
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. SPIE Conf. Ser. Vol. 627, Instrumentation in Astronomy VI. SPIE, Bellingham, p. 733
Van Dokkum P. G., 2001, PASP, 113, 1420
Worley C. C., de Laverny P., Recio-Blanco A., Hill V., Bijaoui A., 2016, A&A, 591, A81
Zalubas R., 1976, J. Res. Natl. Bur. Stand., 80A, 221
Zalubas R., Corliss C. H., 1974, J. Res. Natl. Bur. Stand., 78A, 163

APPENDIX: THXE LINE LISTS

The tables below give the wavelengths of Thorium and Xe lines used in our line lists. Although five decimal places are given for each wavelength, we do not suggest such accuracy. See Section 4.7.1 for the estimation of the precision of the method that 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. Fig. A1 shows a typical arc spectrum with matched arc lines for all four arms.

Table A1. Lines from our blue arm line list and wavelengths from the literature. First column gives the wavelengths we assumed for our line list. 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).

Our λ	Literature λ	Line	Source	Note
		Blue arm		
4715.038 63	4715.18	Xe II	a	†
4723.443 78	4723.4382	Th I	b	
4731.121 27	4731.1955	Th I	b	
4734.152 42	4734.152	Xe I	c	
4740.525 10	4740.5292	Th II	b	
4752.403 23	4752.4141	Th II	b	
4764.346 52	4764.3463	Th I	b	
4769.056 94	4769.05	Xe II	a	
4773.236 34	4773.241	Th I	b	
4775.218 22	4775.18	Xe II	a	
4778.309 90	4778.294	Th I	b	
4779.151 57	4779.18	Xe II	a	
4784.577 00	4784.5667	Th II	d	
4786.516 91	4786.531	Th I	b	
4787.768 40	4787.7462	Th II	d	♣
4789.443 99		Th I, Th II	e, b	♣♣
4792.612 60	4792.619	Xe I	c	
4796.404 38	4796.48	Xe II	a	
4799.326 45	4799.45	Xe II	a	*
4807.028 31	4807.02	Xe I	c	
4808.147 95	4808.1337	Th I	b	
4809.608 32	4809.614	Th I	b	
4817.021 13	4817.0206	Th I	b	
4818.004 78	4818.02	Xe II	a	
4823.274 91	4823.35	Xe II	a	**
4826.675 68	4826.7004	Th I	b	
4829.699 64	4829.71	Xe I	c	
4831.133 37	4831.1213	Th I	b	
4840.883 00	4840.8492	Th I	b	
4843.289 96	4843.29	Xe I	c	
4844.322 27	4844.33	Xe II	a	
4848.341 16	4848.3625	Th I	b	
4850.394 84	4850.23	Xe II	a	*
4853.745 21	4853.77	Xe II	a	
4862.356 82	4862.45	Xe II	a	*
4863.163 31	4863.1724	Th I	b	
4865.472 57	4865.4775	Th I	b	
4872.919 90	4872.9169	Th I	b	
4876.464 65	4876.4949	Th I	b	
4878.743 79	4878.733	Th I	b	
4883.529 98	4883.53	Xe II	a	

Table A1 – *continued*

Our λ	Literature λ	Line	Source	Note
4884.072 36	4884.07	Xe II	a	**
4887.281 88	4887.29	Xe II	a	
4890.056 17	4890.09	Xe II	a	
4894.950 91	4894.9551	Th I	b	
4902.049 18	4902.0545	Th I	b	
		Green arm		
5650.282 89	5650.3276	Th I	b	
5657.945 36	5657.9255	Th I	b	
5659.394 68	5659.38	Xe II	f	
5665.218 02	5665.1799	Th I	b	
5667.555 24	5667.56	Xe II	f	
5670.854 48	5670.91	Xe II	f	
5685.226 15	5685.1921	Th I	b	
5695.810 28	5695.75	Xe I	c	
5696.546 02	5696.477	Xe I	g	
5699.622 86	5699.61	Xe II	f	
5700.965 96	5700.9176	Th II	b	
5707.113 48	5707.1033	Th II	b	
5715.984 84	5716.10	Xe II	f	
5719.623 83	5719.61	Xe II	a	
5720.208 52	5720.1828	Th I	b	
5725.403 98	5725.3885	Th I	b	
5726.913 40	5726.91	Xe II	f	
5741.889 55	5741.82881	Th I	e	
5744.125 09	5744.2	Xe II	a	
5748.763 71	5748.7412	Th I	b	
5749.378 55	5749.3883	Th II	b	
5751.013 75	5751.03	Xe II	f	
5753.086 90	5753.0265	Th I	b	
5758.672 08	5758.65	Xe II	f	
5760.577 06	5760.5508	Th I	b	
5763.578 41	5763.529	Th I	b	
5768.148 54	5768.1812	Th I	b	
5776.391 42	5776.39	Xe II	f	
5789.653 89	5789.6451	Th I	b	
5792.447 89	5792.4304	Th I	b	
5797.478 03	5797.3194	Th I	b	†
5800.843 47	5800.8297	Th I	b	
5804.171 43	5804.1412	Th I	b	
5814.518 30	5814.505	Xe I	g	
5815.342 88	5815.4219	Th II	b	†
5815.927 98	5815.96	Xe II	f	
5823.881 12	5823.89	Xe I	c	
5824.767 30	5824.80	Xe I	c	
5834.993 88	5835.167	Th I	d	†
5856.509 00	5856.509	Xe I	g	
5859.372 51	5859.348	Th I	h	
5868.630 03	5868.374 75	Th I	e	†
5875.018 00	5875.02	Xe I	c	

Table A1 – *continued*

Our λ	Literature λ	Line	Source	Note
	Red arm			
6485.624 36	6485.3759	Th II	b	†
6487.768 93	6487.76	Xe I	c	
6490.761 62	6490.7372	Th I	b	
6494.760 87	6494.981	Th II	d	†
6498.714 37	6498.72	Xe I	c	
6504.222 50	6504.18	Xe I	c	
6506.933 87	6506.9863	Th I	b	
6509.100 27	6509.0503	Th I	b	
6512.783 74	6512.83	Xe II	f	
6521.486 06	6521.508	Xe I	g	
6528.618 11	6528.65	Xe II	f	
6531.356 71	6531.3418	Th I	b	
6533.160 06	6533.16	Xe I	c	
6543.349 82	6543.360	Xe I	g	
6549.352 30	6549.820	Th I	h	†
6551.756 06	6551.7055	Th I	b	
6552.713 40	6552.79	Xe I	j	♣
6554.200 81	6554.196	Xe I	g	
6556.750 37	6556.70	Xe II	a	
6560.040 11	6559.97	Xe I	c	
6563.170 49	6563.19	Xe II	a	
6577.226 45	6577.2146	Th I	b	
6583.888 61	6583.906	Th I	b	
6588.552 75	6588.5396	Th I	b	
6591.483 14	6591.4845	Th I	b	
6593.951 97	6593.9391	Th I	b	
6594.977 30	6595.01	Xe II	f	
6595.563 55	6595.56	Xe I	c	
6597.233 14	6597.25	Xe II	f	
6598.813 54	6598.84	Xe II	f	
6602.758 24	6602.7620	Th I	b	
6619.983 06	6620.02	Xe II	a	
6632.428 26	6632.464	Xe I	g	
6644.653 41	6644.6700	Th I	b	
6657.887 12	6657.92	Xe I	c	
6662.255 22	6662.2686	Th I	b	
6666.947 83	6666.965	Xe I	g	
6668.900 04	6668.92	Xe I	c	
6674.656 33	6674.6969	Th I	b	
6694.259 04	6694.32	Xe II	f	
6694.925 23	6695.181	Th I	d	†
6702.194 61	6702.25	Xe II	a	
6703.853 95	6703.8966	Th II	b	
6719.177 25	6719.199 43	Th I	e	
6728.006 80	6728.01	Xe I	c	

Table A1 – *continued*

Our λ	Literature λ	Line	Source	Note
	IR arm			
7600.764 66	7600.77	Xe I	c	
7618.514 01	7618.57	Xe II	f	
7627.200 55	7627.1749	Th I	b	
7630.267 63	7630.3106	Th I	b	
7636.211 45	7636.174 73	Th I	e	
7642.006 52	7642.02	Xe I	c	
7643.925 43	7643.91	Xe I	c	
7647.407 41	7647.3794	Th I	b	
7653.845 00	7653.8284	Th I	b	
7658.293 78	7658.3202	Th I	b	
7664.587 40	7664.56	Xe I	c	
7666.669 91	7666.61	Xe I	c	
7670.622 63	7670.66	Xe II	f	
7674.414 59	N/A	N/A	N/A	
7678.122 87	7678.1267	Th I	b	
7685.305 65	7685.3075	Th II	b	
7687.748 23	N/A	N/A	N/A	
7698.978 33	N/A	N/A	N/A	
7710.249 28	7710.2689	Th I	b	
7712.249 19	7712.42,7712.405	Th I, Xe II	a	‡
7728.948 25	7728.9509	Th I	b	
7740.351 02	7740.31	Xe I	c	
7783.660 85	7783.66	Xe I	c	
7786.969 75	7787.04	Xe II	f	
7788.949 32	7788.9342	Th I	b	
7798.389 57	7798.357 89	Th I	e	
7802.654 68	7802.65	Xe I	c	
7817.708 70	7817.7669	Th I	b	
7832.979 25	7832.98	Xe I	c	
7837.007 51	7836.703	Th II	k	†
7847.503 51	7847.5394	Th I	b	
7865.973 27	7865.9698	Th I	b	
7881.277 33	7881.32	Xe I	c	
7886.270 84	7886.283	Th I	b	

Notes. †: wavelengths do not 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.

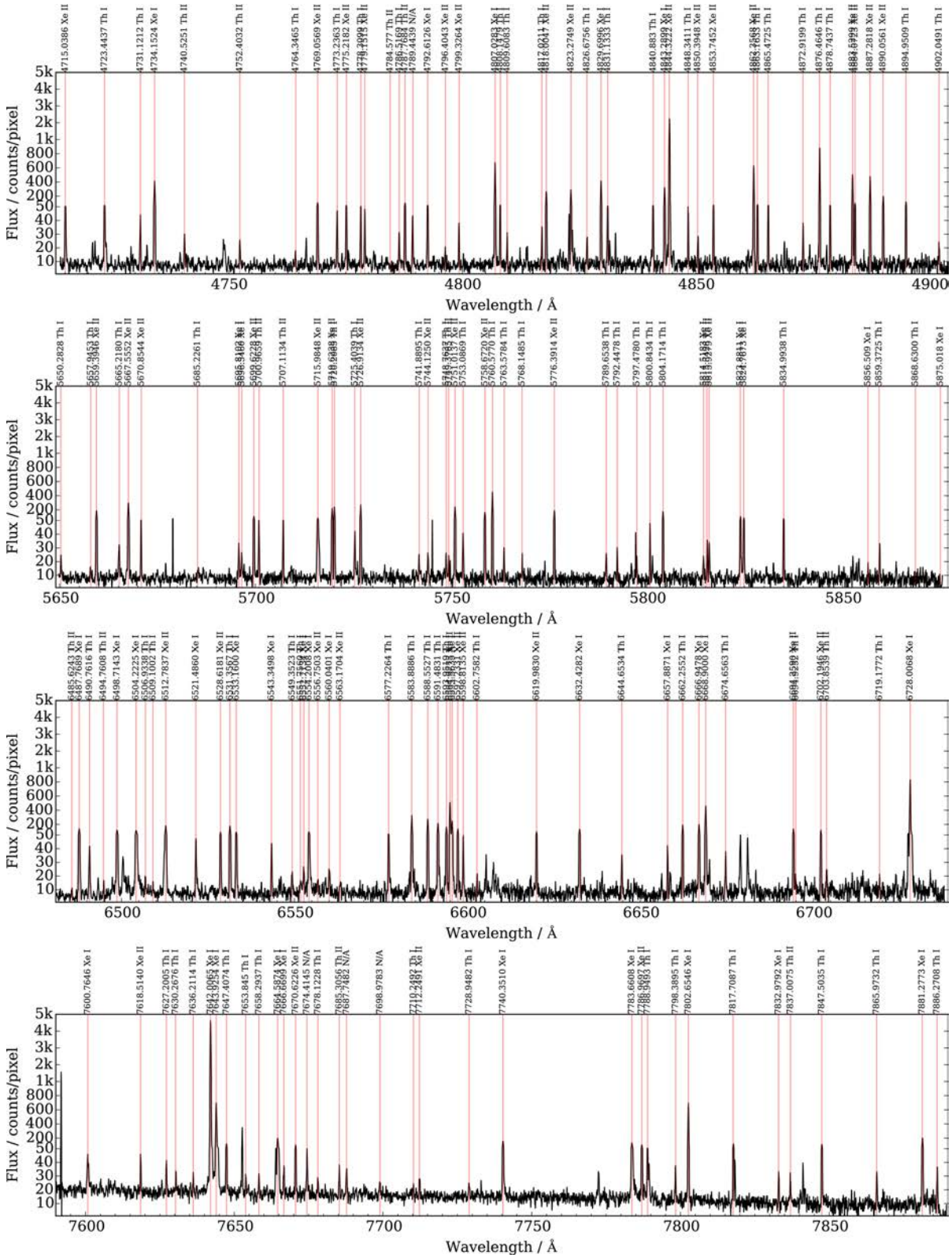

Figure A1. Typical arc spectra with marked lines included in the line lists. Blue, green, red, and IR arms follow top to bottom. Some strong, but unmarked lines are cosmetic errors and some unmarked lines have peculiar shapes, so they did not make it into the line lists. The wavelengths are taken from our line list. There is a lot of dynamic range in the flux, so we used a non-linear flux scale in all four panels.

[1]*Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia*
[2]*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia*
[3]*Research School of Astronomy and Astrophysics, Australian National University, ACT 2611, Australia*
[4]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK,*
[5]*Australian Astronomical Observatory, North Ryde, NSW 2133, Australia*
[6]*School of Physics, UNSW, Sydney, NSW 2052, Australia*
[7]*Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia*
[8]*Research Centre in Astronomy, Astrophysics and Astrophotonics, Macquarie University, Sydney, NSW 2109, Australia*
[9]*Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*
[10]*Sydney Astrophotonic Instrumentation Labs, School of Physics, A28, The University of Sydney, NSW 2006, Australia*
[11]*Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia*
[12]*International Centre for Radio Astronomy Research (ICRAR), The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*
[13]*INAF National Institute of Astrophysics, Astronomical Observatory of Padova, I-36012 Asiago, Itay*
[14]*Australian Centre for Astrobiology, UNSW Australia, Sydney, NSW 2052, Australia*

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