HiFLEx—A Highly Flexible Package to Reduce Cross-dispersed Echelle Spectra

Ronny Errmann[1] , Neil Cook[2] , Guillem Anglada-Escudé[3], Sirinrat Sithajan[4], David Mkrtichian[4] , Eugene Semenko[4],
William Martin[1] , Tabassum S. Tanvir[1], Fabo Feng[1,5] , James L. Collett[1], and Hugh R. A. Jones[1]

[1] The Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK
[2] Institut de Recherche sur les Exoplanètes, Département de Physique, Université de Montréal, Montréal, QC H3C 3J7, Canada
[3] Institut de Ciències de l'Espai Barcelona, E-08035, Spain
[4] National Astronomical Research Institute of Thailand (NARIT), 260 Moo 4, T. Donkaew, A. Maerim, Chiangmai, 50180, Thailand
[5] Earth & Planets Laboratory, Carnegie Institution for Science, 5241 Broad Branch Road, NW, Washington, DC 20015, USA

Abstract

We describe a flexible data reduction package for high resolution cross-dispersed echelle data. This open-source package is developed in `Python` and includes optional GUIs for most of the steps. It does not require any pre-knowledge about the form or position of the echelle-orders. It has been tested on cross-dispersed echelle spectrographs between 13k and 115k resolution (bifurcated fiber-fed spectrogaph ESO-HARPS and single fiber-fed spectrograph TNT-MRES). HiFLEx can be used to determine radial velocities and is designed to use the TERRA package but can also control the radial velocity packages such as CERES and SERVAL to perform the radial velocity analysis. Tests on HARPS data indicates radial velocities results within $\pm 3\,\mathrm{m\,s}^{-1}$ of the literature pipelines without any fine tuning of extraction parameters.

Key words: Radial velocity – Spectroscopy – Astronomy – data reduction

Online material: color figures

1. Introduction

High resolution echelle spectrographs are a highly important tool to study astrophysical phenomena like the study of stellar atmospheres (e.g., Ryde et al. 2019) or the search for and characterization of extrasolar planets around nearby stars (e.g., Anglada-Escudé et al. 2016; Lovis et al. 2017). High resolution is reached by using an echelle grating, which is optimized for reflection in high orders, as higher orders have higher dispersion. However, these orders are overlapping in space, therefore a cross disperser is used to separate the orders with dispersion perpendicular to the echelle dispersion.

For precise results, instruments require precise data reduction packages. Many of the current high resolution spectrographs have dedicated data reduction packages, e.g., DRS (Data Reduction System) for High Accuracy Radial velocity Planet Searcher (HARPS) and ESPRESSO (Sosnowska et al. 2015; Di Marcantonio et al. 2018), CARACAL for CAR-MENES (Caballero et al. 2016), or APERO for SPIRou/NIRPS (N. Cook et al. 2020, in preparation). However, for some echelle spectrographs no data reduction pipeline is available.

Additional to the data reduction packages provided for a single instrument, more general data reduction packages which can be applied to a larger set of echelle instruments are available. Packages to extract data from stabilized spectrographs are available (e.g., Zechmeister et al. 2014), however, only some echelle spectrographs are stabilized. Another example is the CERES pipeline (Brahm et al. 2017), which allows the reduction of data from HARPS (Mayor et al. 2003), FEROS (Kaufer et al. 1999), HIRES (Vogt et al. 1994), CORALIE, and several other spectrographs. However, for each instrument the pipeline requires a bespoke script with several hundred lines of code and prepared configuration files with the positions of the orders and the wavelength solution. This makes for quite a lot of extra steps to apply it to data from spectrographs not included. The preset requirements make data reduction for an instrument in development, when the position of the echelle orders or resolution might change often, a very considerable overhead.

We are developing a new high resolution spectrograph (Lhospice et al. 2019) for the Thai National Telescope (TNT). The spectrograph will be fiber-fed with a bifurcated fiber input and a foreseen resolution of 100,000. We test a variety of different designs. The available pipelines were not sufficiently flexible to enable quick analysis of our data as many variations in the code were required for each new setup. Instead we

developed a flexible data reduction package, in which a single calibration file has to be edited while no modification in the program code are required to reduce data from a modified setup. This flexible package also allows the reduction from data of a wide range of echelle spectrographs. We tested the package on data from established spectrographs.

After the extraction and calibration of the spectrum, precise radial velocities can be measured using available packages: TERRA (Anglada-Escudé & Paul Butler 2012), SERVAL[5] (Zechmeister et al. 2018), and CERES pipeline.[6] The description of these packages can be found in the corresponding publications. Further radial velocity (RV) packages are available to analyze the reduced spectra, e.g., Bedell et al. (2019), Rajpaul et al. (2020). Increase precision of the barycentric corrections can be done with packages like PEXO (Feng et al. 2019).

2. Description of the Package

The purpose of the HiFLEx pipeline is to reduce and analyze echelle spectra with minimal human interaction. The pipeline works for echelle spectrographs with either single fiber or bifurcated fiber input if at least two distinguishable echelle orders are imaged onto the detector. The data reduction package is written in `Python` and has a modular setup. The `numpy` and `scipy` packages of `Python` are implemented for most calculations. The used fitting routines are based on the least square minimization Levenberg-Marquardt algorithm (for polynomial or Gaussian fits) or the `dgelsd` routine which uses Singular value decomposition (for two-dimensional polynomial fits). HiFLEx can be downloaded from Github.[7] In this section we describe the essential steps carried out by the pipeline, for the practical application we refer to the manual.[8] The flow chart of the package is described in Figure 1 and will be described in this chapter.

2.1. Preparation

All configuration can be done in a configuration file. A description of the available parameters can be found in the manual provided with the package and in the configuration file itself. In Table 1 we provide a list with the most important parameters to be modified to set up the extraction of a new spectrograph.

The package assigns the available frames to the calibration steps depending on the header of the files and the name of the file using a common naming convention, which is given in the manual. Afterwards the pipeline allows the user to modify

5   https://github.com/mzechmeister/serval
6   https://github.com/rabrahm/ceres
7   https://github.com/ronnyerrmann/HiFLEx/
8   https://github.com/ronnyerrmann/HiFLEx/blob/master/HiFLEx_UserManual.pdf
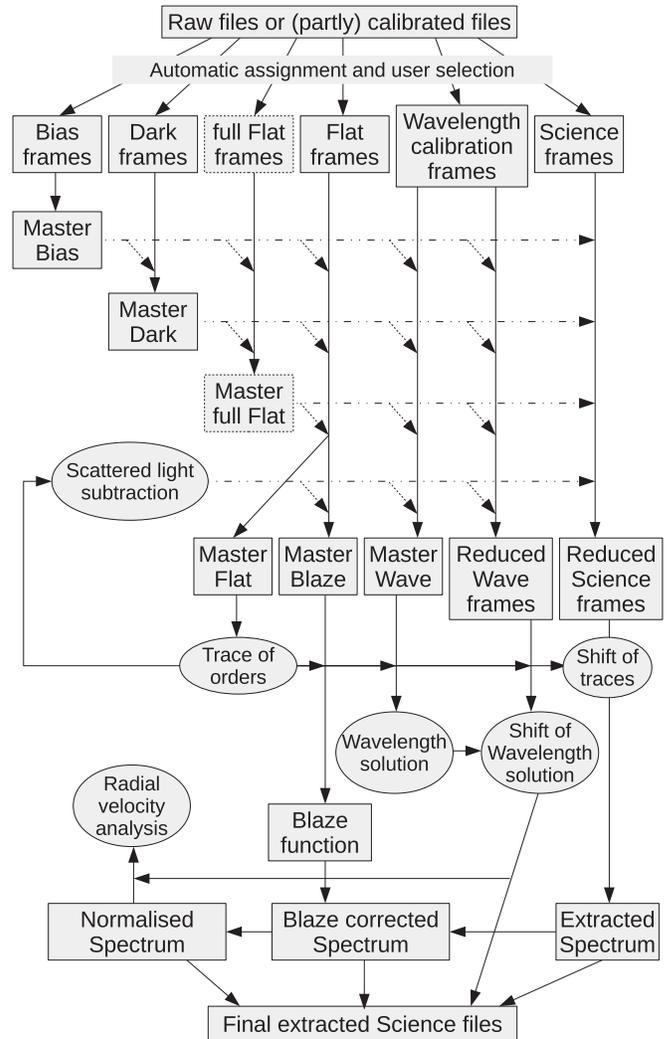
Figure 1. Flow chart of the data reduction and analysis of HiFLEx. The chart shows all available steps for data processing and a suggested usage of calibration data. Wave stands for the images from which a wavelength solution can be determined, e.g., an emission lamp spectrum. However, the pipeline also works with subsets of these steps. For example, taking full frame flats (the detector is illuminated uniformly) is not possible for most spectrographs, hence this step is most often skipped. On the other hand, if only a quick glance of an uncalibrated (no wavelength solution) spectrum is required, then a single science frame is enough, which can be used as master flat and master blaze. Everything between this extreme and full calibration is possible.

the assignment of the individual observation images for the different calibration steps.

2.2. Image Processing and Master Files

The calibration of the image frames depends on the instrument. For example, a CCD detector cooled to liquid nitrogen temperatures might not require dark frame correction when observing bright targets, while dark correction is highly

Table 1
List of the Most Important Parameters from the Configuration file to Set up HiFLEx for a new Spectrograph

Parameter	Short Description
GUI	*True* (modify once appropriate setup has been found)
raw_data_paths	Folders of the raw data
rotate_frame, flip_frame	To orient the raw images: vertical traces, reddest pixel at (0,0)
arcshift_side	Calibration traces are *left* to, *right* to, or at the same (*center*) position as science traces
original_master_wavelensolution_filename	Set to new file to use a GUI to create new wavelength solution (or set to previous solution)
raw_data_*	Set the keyword names and values to assign files to reduction steps automatically
*_calibs_create_g	Define the processing of the raw files for each image type or reduction step
altitude, latitude, longitude	Observation site, if not in header
terra_jar_file, path_ceres, path_serval	Paths to external radial velocity packages (optional, need to be downloaded manually)

Note. Some suggested default values are given in italic. Detailed information can be found in the manual provided with the pipeline (see footnote 7). In order to run the pipeline, "run file_assignment.py" (to check and modify assignments of raw files to calibration files or reduction steps) and then "run hiflex.py" (to perform the calibration steps and extract the science spectra).

recommended for a Peltier cooled frame from a CMOS architecture detector.

The HiFLEx package allows the user to apply a broad variety of calibration steps. Additionally, pre-reduced images or master images can be given, to which none or only some calibration steps will be applied. Possible calibrations include windowing, bad-pixel mask correction for dead pixels, correction with dark, bias, or flat field images, and removal of background (scattered) light measured between the orders.

If the user provides a bad-pixel mask, these bad pixels will be replaced by the median value of the surrounding pixel, using a 3×3 box around the bad pixel. Extracted spectral regions that contained bad pixels during the extraction will be marked so they can be excluded from further analysis.

The master files (e.g., Bias, Dark, Flat-field, or image to trace the orders or extract the blaze function) are created by the pipeline once they are required using the (raw-) frames and calibration steps as given by the user. The images are median-combined on a pixel-by-pixel basis. However, the user can also combine images using the sum or average. Before any combination, images can also be scaled to the average median flux of the frames before combining, e.g., if the light source varies in time. Master frames already created by the pipeline are automatically used or those specified by the user will be read.

2.3. Finding and Defining the Traces of the Echelle Orders

While it is a relatively easy task for humans to define the traces of echelle spectra, this task is harder to program in a general way. For example in some spectrographs the red orders are partly overlapping or the blue orders have a much smaller signal than the stray light between the red orders. To keep flexibility in our package, the ability to trace the orders should not depend on the curvature of the orders. For this step the user can use any data, however, we suggest to use a master file created from several high signal-to-noise continuum spectra (e.g., Tungsten) in the science fiber. In the following explanation we assume that the traces of the orders are vertically oriented.

Our approach for finding the traces is two-step. At first, in a binned image the traces are searched. Binning increases the signal-to-noise ratio and decreases the calculation time. In tests we saw that using the mean instead of the median for binning allows a better tracing, especially when large bin sizes in the cross-dispersion direction can be used due to large gaps between the traces of the orders. Second, the iterative finding of all orders is done. The pixel with the highest flux is searched and a Gaussian function fitted to the one-dimensional cross section in the cross-dispersion (horizontal) direction (with a width 1 pixel in the vertical direction). The center of the

Gaussian fit is assumed to be the center of the trace at this dispersion pixel. The Gaussian fit is then repeated on one-dimensional cross section above and below the previous position, using the previous fitted center as the starting point until the fit fails on several consecutive cross sections, e.g., because the Gaussian fit is less significant than 3σ. If the information from other traces is available, then this information is used to follow the curvature of the trace in areas where no fit is possible. Once the trace has been identified, a polynomial fit from user-defined polynomial degree N_{poly} (default 3) is fitted to the center of the Gaussians $\mathrm{px}_{\mathrm{cen}}$ for each pixel in the trace px_y:

$$\mathrm{px}_{\mathrm{cen}} = \sum_{i=0}^{N_{\mathrm{poly}}} a(i)(\mathrm{px}_y)^i, \qquad (1)$$

with $a(i)$ as the coefficients of the polynomial. The lowest polynomial degree to fit the data should be chosen for N_{poly} to avoid over-fitting. The results can be checked in the GUI or the logged images. Traces will only be used, if they were identified in more than 25% of the image and if they do not cross previously found traces. The algorithm has been tested in images with cosmic rays and ghosts, as well as in images with high scattered light between the traces. The identification is robust against these effects. After one order has been traced, the trace of the order is masked in the binned image and the next brightest pixel searched to repeat the above steps.

The second step consists of redefining the traces and finding the extraction width of the traces. This step can be done on an image that is binned slightly in dispersion direction to speed up the process. For the previously found traces the Gaussian fit in cross-dispersion direction is repeated. The brightest area of the cross-section is also fitted with a third order polynomial to define the center of the trace for non-Gaussian profiles. Additionally the borders of the trace, where the flux drops to a user-defined level (default 5%) are measured by applying a linear regression to the transition area. These measurements are repeated for each cross section of the trace in dispersion (vertical) direction. The final trace and the extraction borders are defined by a polynomial fit of user-defined polynomial degree (default 3) along the center of the Gaussian and along the measured position of the borders for each side. Figure 3 shows an image with the marked traces and Figure 2 a zoom showing the extraction borders (red-dashed lines).

For the case of a bifurcated fiber feed, the traces of the calibration fiber will be determined by shifting the position of the previously identified traces in images in which the calibration fiber is illuminated. The shift between the traces of the science order and its corresponding calibration order depends on the dispersion of the cross-disperser (especially when a prism is used). For resilience against badly identified shifts which can happen when orders have very low signal, a second order polynomial is fitted to the shift of each calibration
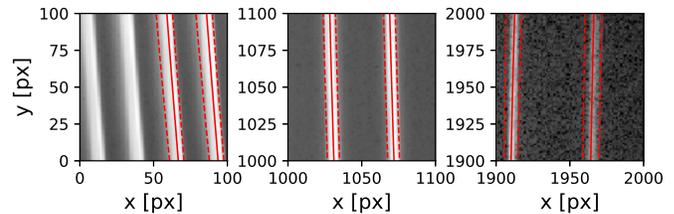
Figure 2. HiFLEx identified orders in three different areas of the detector: two opposite corners (left and right panel) and the center (middle panel). The data was taken with the echelle spectrograph in development for the TNT. The center of the order is marked with a solid line, the extraction area (the flux is above 5% of the maximum flux) with dashed lines. Two orders in the left panel are not identified, as only a small portion of the orders are presented in the image. The whole arrangement of orders is shown in Figure 3 with the same pixel labels.
(A color version of this figure is available in the online journal.)

trace (s_o):

$$s_o = \sum_{i=0}^{2} a(i)o^i, \qquad (2)$$

with o as order of the trace. Sigma-clipping is applied to remove outliers from the fit (default 3σ). The final shifts are then derived from Equation (2). These shifts are added to the constant term in the trace of each order, while the higher order polynomial terms are kept the same to preserve the curvature of the traces.

2.4. Wavelength Solution

Once the position of the traces is known, a spectrum for the wavelength calibration can be extracted. The positions of these lines will then be compared with a standard list of wavelengths to find the relation between pixel-position and the wavelengths. The pipeline will use a user-defined line list for the reference wavelengths. A line list for Th–Ar and U–Ne is provided. Both lists were taken from the NIST database[9] (Kramida et al. 2018). Any other line list can be provided by the user. For example, if a Fabry–Perot interferometer or laser frequency combs is used to perform the wavelength calibration. The usage of an equally spaced wavelength source requires knowledge of the approximate wavelength calibration within a few pixel beforehand, otherwise the line list can be used as any other line list from emission lamps.

For best precision, the user should limit the line list to stable lines and the line list should be cleared of blended and asymmetric lines to avoid these lines affecting the precision of the wavelength solution. For example in the case of a spectrograph that covers the spectral range of HARPS, the HARPS catalog can be used (Lovis & Pepe 2007). For naming

---

[9] https://physics.nist.gov/PhysRefData/ASD/lines_form.html

convention, we will assume from here on an emission line spectrum from a Th–Ar lamp.

The dispersion θ_m of light of wavelength λ is defined by the grating used and follows Equation (3), with the spacing d between the groves of the grating, order number m, and incident beam angle θ_i:

$$d(\sin\theta_i - \sin\theta_m) = m\lambda. \quad (3)$$

For a single order (dispersion direction) Equation (3) leads to $\lambda \propto \sin\theta_m$, which can be approximated by a low-order polynomial to correlate pixel and wavelengths. Using a low-order polynomial also corrects distortions that might be introduced by the lens system for off-axis beams or other effects introduced by the optical system. In case of constant θ_m (same pixel in each order, cross-dispersion direction) the wavelength λ in each order follows $\lambda \propto \frac{1}{m}$ (Equation (3)). However, the lens and cross-disperser might cause deviation from this equation due to their chromaticity, hence a higher-order polynomial might be necessary to correlate the order number and central wavelength. Finally, the correlation between inverse number of the order $\frac{1}{m}$, pixel-position px, and wavelength λ will be done with a two-dimensional polynomial:

$$\lambda(\mathrm{px}, m) = \sum_{i=0}^{N_{\mathrm{poly,C}}} \sum_{j=0}^{N_{\mathrm{poly,D}}} a(i,j)\left(\frac{1}{m}\right)^i \mathrm{px}^j, \quad (4)$$

with $a(i,j)$ as coefficients of the polynomial. The user can define the degree of the polynomial along the order-number ($N_{\mathrm{poly,C}}$) and the pixel-position ($N_{\mathrm{poly,D}}$).

The position of the emission lines in the extracted spectrum are searched in a similar way as the position of the traces. For each order a high order polynomial is fitted to the extracted flux. At all locations with significant remaining flux, the precise center together with other line properties is determined by fitting a Gaussian function. A list with all the emission line positions, widths, and heights is created in this way. This list can be cleared by the user to allow removal of blended lines.

In the next step the found lines have to be compared with the set of reference lines for the corresponding lamp. If a suitable wavelength solution exists, the pipeline continues with the step described in Section 2.4.2.

2.4.1. Creation of a New Wavelength Solution

The first time a completely new setup or instrument is used the user has to provide a file with order, pixel, and wavelength as an initial guess for the wavelength solution. This step can be done in a GUI that shows the list of found emission lines, as well as the visualization of the extracted (Th–Ar) spectrum. The input will be fit with a user-defined two-dimensional polynomial (Equation (4), default 3 for both dimensions). A suitable solution is then improved.

2.4.2. Adjusting from a Previous Wavelength Solution

For most spectrographs the wavelength solution between individual nights will be nearly constant with only small changes. For stabilized spectrographs the variation will be in the sub-pixel regime. However, for the purpose of spectrograph testing the pipeline provides the user with the possibility to adjust a previous wavelength solution with a wide set of options. In case a different number of orders are identifiable between different nights, a user-defined offset range for physical dispersion order can be given. A search for a shift in dispersion direction and a variation of resolution can be performed too.

For each step in these multidimensional variations the number of correlated emission lines between the spectrum of the calibration frame and the catalog is calculated. For the best matching settings, the new wavelength solution is created by fitting the correlation between inverse number of the order $\frac{1}{m}$, pixel-position px, and wavelength λ with a two-dimensional polynomial (Eqation (4)) of user-defined polynomial degree (default 3 for both dimensions). The step is done iteratively, removing the least matching correlation between emission lines in the spectrum and reference catalog. Figure 3 shows an image with the marked emission lines.

2.5. Drift of the Wavelength solution

For spectrographs that do not use simultaneous wavelength monitoring (like Th–Ar or an iodine cell) the drift of the spectrum can be measured only before and after an science exposure (or less often to avoid the loss of science time). Dealing with the drift is discussed in Barnes et al. (2014). HiFLEx provides the option to use calibration spectra to correct any drift of the spectrum. In this case the wavelength solution will be interpolated using measurements closest before and after the exposure.

For bifurcated-fiber spectrographs the wavelength solution can be adjusted using the simultaneous Th–Ar light from the calibration fiber. Furthermore, the shift in the wavelength solutions between science and calibration fiber can be determined if images with Th–Ar in both fibers are available. The wavelength solution will be determined for both fibers and the coefficients in Equation (4). Adjusted for each solution and the shift determined from the differences. This shift is interpolated using the measurement adjacent to the science exposure and applied to the wavelength solution of the science spectrum.

2.6. Blaze Function

The efficiency of the echelle grating varies along each order, with the highest efficiency in the center of the order and decreasing efficiency at the corners. This is called the blaze function. Additionally, the throughput of the whole optical
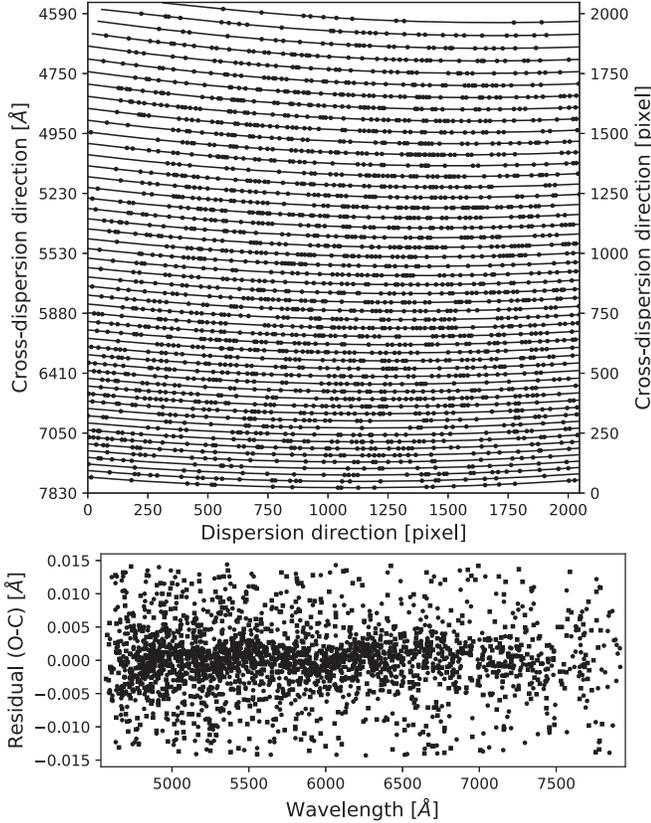
Figure 3. Top: traces of the orders and position of the Th–Ar lines used for the wavelength solution. Bottom: residuals (Observed—Calculated) from the fit of the wavelength solution in Angstrom, shown for different wavelengths. The data was taken with the prototype of an echelle spectrograph.

system depends on the wavelength. In order to correct this effect a high signal-to-noise spectrum of a standard star or a continuum source can be used. HiFLEx allows the use of an extracted continuum spectrum (e.g., for stabilized spectrographs) or fits a polynomial along the extracted flux in each order to obtain the blaze function. Prior to fitting, the spectral features are sigma-clipped (default 4σ).

2.7. Extracting Science Spectra

Before extracting the science spectra, the user-defined data reduction such as dark, bias, or background correction are performed. The background correction allows removal of stray light by measuring the signal between the orders and fitting a low order two-dimensional polynomial. HiFLEx can also account for any drift of the spectrum in the cross-dispersion direction in the next step, where all traces are shifted to find the position in which the highest flux is extracted. To speed up the process, only the light of the central column of pixels of the shifted traces is used. To find the best shift, a Gaussian function is fitted to the total extracted flux of each shift.

The light from the science fiber is extracted from the reduced image using the shifted traces. The light for each dispersion step pixel is summed up. For the extraction of the flux at the borders of the traces the user has the option to either use the fractional flux of the pixel or the interpolated flux. The interpolated flux will produce more precise results, however it requires more computational time. Furthermore, if the border of the extraction area is at the level where the light drops down to only few percent of the maximum of this dispersion step, the improvements are typically less than 1 ppm.

In the case of simultaneous observation of an emission line spectrum in the calibration fiber, this spectrum is extracted using the same shift as determined for the science traces. This spectrum is then used to determine the shift between the emission lines and the wavelength solution by fitting a Gaussian function to all lines used to define the wavelength solution. The master shift of the wavelength solution is calculated from the median of the re-identified emission lines.

In case of well exposed orders, the uncertainties of the extracted flux are dominated by the photon noise ($\sqrt{\text{flux}}$). For orders with low flux-levels additional noise sources become important. This includes the readout noise, the noise from stray light, and the noise from the dark. HiFLEx uses the noise measured between the traces of the echelle orders to estimate this uncertainty. The value for the uncertainties for the extracted data is calculated using

$$\sigma_{\text{flux}} = \sqrt{\text{flux} + n_{\text{pixel}}\sigma^2_{\text{background}}},$$

with the extraction width of the order denoted by n_{pixel} and $\sigma_{\text{background}}$ being the standard deviation of the signal in the image with the traces masked out.

In the next step a blaze corrected spectrum is derived by dividing the extracted spectrum with the blaze function (see Section 2.6). The blaze function is normalized to a median flux of one before the division. The uncertainties for the blaze corrected spectrum are calculated from the variation of the residuals when fitting a high order polynomial against the windowed spectrum.

The continuum of the blaze-corrected spectrum is normalized by fitting a low order polynomial to the spectrum of each order. Absorption and emission lines are masked beforehand. The noise in the spectra is measured along the continuum parts of the spectrum by fitting a low order polynomial along the measured noise. To remove the signal introduced by cosmic rays an optional sigma-clipping can be performed using a 4σ threshold, while excluding the areas of absorption lines for the calculation of the scatter.

2.8. Storing the Data

To provide compatibility with other software packages, the reduced data is automatically stored in a variety of different ways. The pipeline stores data in the same way as the CERES

Figure 4. Images of a long-exposed Th–Ar spectrum taken with MRES. The brightness is scaled to logarithmic. The traces of the orders are shown (alternating red, blue, green).
(A color version of this figure is available in the online journal.)

pipeline (Brahm et al. 2017), in the format provided by ESO DRS,[10] as single files with IRAF compatible headers, and in a csv-file for the use with TERRA. The manual provides further information. If coordinates are provided by the user, the pipeline calculates barycentric correction for time and velocity (Kanodia & Wright 2018) and stores the information together with other results and statistics in the image header.

3. Results

We have tested the package with a variety of different sets of data. This includes data taken with different detectors and different setups of a prototype spectrograph (H. R. A. Jones et al. 2020, in preparation). The package was able to trace the orders in a variety of challenging setups, which also included orders of high curvature and partly defocused orders. Furthermore, data taken with established echelle spectrographs has been analyzed, e.g., HARPS and Medium Resolution Echelle Spectrograph (MRES at the Thai National Telescope, Buisset et al. 2018). MRES does not have a data reduction package whereas HARPS data can be analyzed with different packages. The resolution of the respective spectrographs is 13,000 and 115,000.

3.1. Reduction of MRES Spectra

We observed stars which might have had or in the future will have close encounters with the Solar System (Tanvir 2018, T. S. Tanvir et al. 2020, in preparation) with MRES. Possible close encounters of the solar system have been found by modeling the orbits of 14,374 stars using the Gaia DR2 data. MRES spectra provide a better RV determination than the Gaia DR2 results and hence allow more precise orbits. We observed some of the most likely targets from that sample between 2018 November and 2019 April (programme ID C06_019). The observation included

taking all necessary calibration data (Tungsten spectra, Th–Ar spectra with different exposure times, biases). MRES is not a stabilized instrument, hence a few times during the night Th–Ar spectra were taken to monitor the drift of the spectrograph. The detector was cooled to $-100\,°C$. The exposure time for the science spectra were between 45 and 60 minutes, with each target being observed once.

The data was reduced using the HiFLEx pipeline with the parameters described in the manual. Only bias correction was applied to the raw images. Figure 4 shows the Th–Ar spectrum of MRES. The marked traces were identified in the master image of the Tungsten spectra. The red echelle orders have a small separation between each other due to limited diffraction by the cross-dispersing prism. However, even the slightly overlapping orders could be identified by HiFLEx.

After the preparation for the science extraction have been finished, the spectra of the target stars were extracted. Figure 5 shows the spectrum of one of our science targets (first entry in Table 2). The figure shows wavelength ranges with larger amounts of deep absorption lines, for example the telluric lines around 7650 Å. In a close-up of the overlapping orders (bottom panel in Figure 5), lower flux in the blue side of the orange order and therefore lower signal-to-noise ratio is visible. That is the spectra from blue edge of an order (and the red spectrum from red side of another order) have lower signal-to-noise compared to the green spectra because grating efficiency falls away significantly away from center of echelle orders. However, the normalization is not affected by this. The final radial velocities were measured only with the RV code of the CERES package, as this code provides absolute RV against an intrinsic template spectrum. Table 2 gives the RV values determined in this way. Our results are consistent with the radial velocities from other sources for most stars and only in a few cases, e.g., for the targets 5295..., 2946... and 3376... do the MRES velocities indicate significantly different velocities to

Table 2
Radial Velocity (RV) Measurements for the Most Probable Close Encounters of the Solar System

Gaia DR2 Designation	G (mag)	d_{nom} (pc)	t_{nom} (Myr)	v_{range} (km/s)	Gaia RV	HiFLEx RV	LAMOST-DR4 RAVES5 RV
510911618569239040	8.9	0.43	−2.86	25.9-27.1	26.45 (0.35)	29.93 (0.28)	
436664033586936704	11.1	0.36	−2.13	−28.0-29.2	−28.59 (0.35)	−29.33 (0.29)	−28.94 (0.63)
29068050080489144944	10.7	0.42	−1.71	43.6-51.7	47.55 (2.37)	45.95 (0.29)	48.45 (1.53)
5571232118090082816	11.8	0.18	−1.17	81.6-83.1	82.18 (0.47)	88.08 (0.30)	
5700273723303646464	12.0	0.18	−1.64	36.5-39.6	38.05 (0.91)	37.32 (0.29)	
5599691155509093120	12.2	0.33	−0.84	71.3-78.5	75.01 (2.19)	72.15 (0.46)	
52952724810126208	12.2	0.26	−0.54	32.8-43.5	37.77 (3.35)	30.55 (0.79)	25.70 (6.05)
2946037094755244800	12.3	0.24	−0.91	36.9-47.4	42.11 (3.20)	29.21 (0.44)	
955098506408767360	12.4	0.09	−0.74	34.9-42.0	38.52 (2.12)	39.04 (0.77)	33.19 (6.31)
4472507190884080000	12.9	0.36	1.81	−26.7-77.7	−52.18 (15.23)	−57.56 (0.58)	
3376241909848155520	12.5	0.48	−0.45	70.3-89.2	79.91 (5.63)	67.27 (0.63)	

Note. The measurements are based on Gaia DR2 (Gaia Collaboration et al. 2018), HiFLEx, LAMOST-DR4 (http://dr4.lamost.org) and RAVE5 (Kunder et al. 2017). The encounters are based on the range of probable velocities (5–95%) confidence interval, (v_{range}) for a nominal closest approach distance (d_{nom}) at a nominal time (t_{nom}) where negatives values are past encounters. All RVs are given in km s^{-1}, uncertainties are given in brackets.
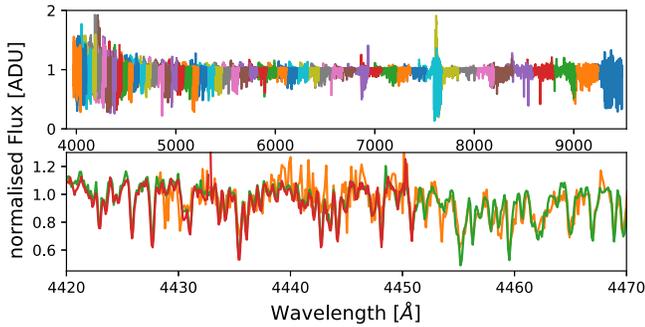
Figure 5. Spectrum of Gaia DR2 510911618569239040, one of the target stars, for MRES after blaze correction and normalization of the spectrum to a continuum flux of 1 ADU. The top panel shows the light from all 52 MRES orders, with each order being in one color. The bottom panel shows the overlap between several orders around 4450 Å. The spectral region plotted is fully covered by the order in green, whereas a bluer order plotted in red finishes around 4450 Å and a redder order in orange starts around 4430 Å.
(A color version of this figure is available in the online journal.)

indicate substantially different orbits making it appropriate to obtain further follow-up observations at different epochs to further investigate the orbits of these stars.

3.2. Comparison of HiFLEx Extraction to Other Packages

In order to more systematically benchmark the package, data analysis has been performed on publicly available data from HARPS. Here we focus on spectra taken of the asteroid Ceres and the corresponding calibration to compare results between other data reduction and RV analysis packages. The data of Ceres provides a reflected solar spectrum. The data were taken in one night (2015 July 30) with HARPS at the 3.6 m telescope at ESO's La Silla observatory.
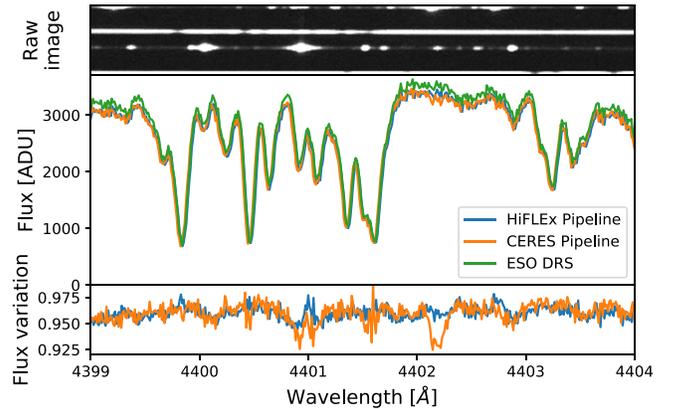
Figure 6. Example of the extracted data for the asteroid Ceres taken with HARPS. The top panel shows the raw data of order 23 of the blue chip for both the science and calibration fiber and the next closest traces. The middle panel shows the extracted spectrum of the science fiber using different packages. The ESO DRS results have been converted from e^- to ADU. The CERES pipeline provides the wavelength in vacuum reference format and with barycentric correction, these two corrections were undone in order to compare with the ESO DRS wavelengths. The bottom panel shows the difference between the HiFLEx and CERES extracted spectrum, when these spectra are divided by the ESO DRS spectrum.
(A color version of this figure is available in the online journal.)

For the different extraction packages tested, an area of the extracted spectrum is shown in Figure 6 for the chosen regime in the center of a central order on the blue chip. The variation of the normalized spectrum relative to the ESO DRS is 0.6% for HiFLEx and 1.0% for CERES. This variation appears to have its origin in the width of the traces. In the HiFLEx configuration file the area to be extracted was set to where the light in the image of the flat field lamp spectrum is above 5% (see footnote 8). This means that some light (with the lowest signal-to-noise ratio) outside of the area is not extracted.
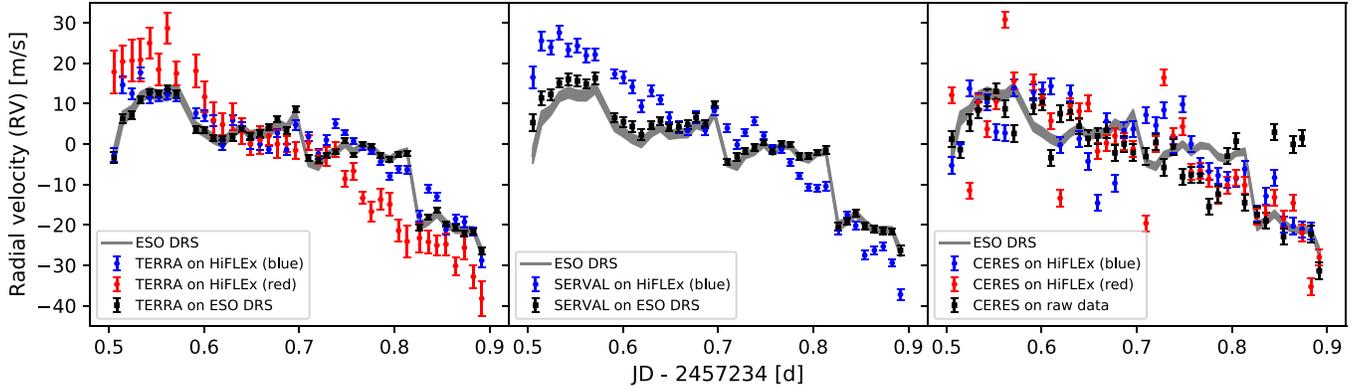
Figure 7. Relative radial velocities of HARPS spectra of the asteroid Ceres using different packages. Each panel in turn shows the results from one package: TERRA, SERVAL, CERES. In gray the results provided by the ESO DRS are given for comparison. HiFLEx was applied to each of the two HARPS CCD individually, blue and red referring to the wavelength range of the CCD (circles). TERRA and SERVAL were applied to the extracted spectra (e2ds-files) from ESO DRS (black squares). CERES was used to reduce the raw-data and measure the RVs in this spectra (black squares). SERVAL results are not shown for red CCD.
(A color version of this figure is available in the online journal.)

However, other data reduction packages use other thresholds leading to small variation in the flux levels. Iterative change of the parameters could be done to optimize and achieve a spectrum closer to the fixed ESO DRS pipeline. For example the curvature of the orders could be defined with a higher order polynomial or the extraction width could be changed. The largest difference between the packages appears to occur in the vicinity of bright emission lines (see top panel of Figure 6). HiFLEx does not correct for small scaled stray light, while the other two packages correct for this in different ways. This may account for the apparent small features at 4401 Å and 4402.2 Å in the CERES spectrum where it can be seen that the flux is few percent lower.

The RV of the data was measured with the packages TERRA, SERVAL, and CERES. TERRA and SERVAL build a master spectrum from the data and measure the radial velocities against this spectrum. However, the barycentric velocity varied about $1 \, \mathrm{km \, s^{-1}}$ in the observation window. This means that unmasked telluric lines might not be averaged out and hence might affect the the templates built by TERRA and SERVAL. This effects the red HARPS chip in particular. On the other hand, the CERES pipeline measures the RV using template spectra. It re-scales the error bars using parameters determined for each instrument by computing the RV of a high S/N spectrum when different noise levels are added (Brahm et al. 2017). In this work, we have used the same parameters as provided by the CERES pipeline for both HARPS chips together. However, as our measurements were based on the individual chips and therefore less orders were available to determine the radial velocities, the error bars are underestimated. Furthermore, this leads to similar error bars for the blue and red chip.

Figure 7 shows the RV measurements analyzing the data of asteroid Ceres with different packages. In the HiFLEx extraction, the data from the two CCD chips in HARPS were analyzed

individually. The RV precision is lower and scatter higher, compared to packages analyzing all orders from both chips at once. This is particularly the case for the red chip as the number of orders and the number of stellar absorption lines per order is lower and can be seen in the much larger error bars in the TERRA RVs. For SERVAL we excluded the RVs from the red chip completely. The RV analysis using CERES shows a higher scatter compared to the results from the other two packages. This is seen when reducing the data with CERES (black squares) and as well when determining the RV of the extracted HiFLEx spectra for each chip individually (blue and red circles).

In this comparison the extraction parameters were not optimized to the HARPS spectrograph. The flexibility of the HiFLEx package has an impact on the precision of its wavelength solution. This apparently arises because our automatic process requires that only Th–Ar lines above a 2σ-threshold are used and lines that deviate from the solution are clipped.

We want to emphasise that this comparison is very limited in scope. Figure 7 serves to illustrate that HiFLEx can obtain RVs within $\pm 3 \, \mathrm{m \, s^{-1}}$ (based on the blue chip) of sophisticated bespoke packages (ESO DRS) without any particular fine tuning of extraction parameters when using TERRA, $7 \, \mathrm{m \, s^{-1}}$ when using SERVAL, and $6 \, \mathrm{m \, s^{-1}}$ when using CERES.

4. Conclusion

We present the data reduction package HiFLEx that allows the extraction and analysis of cross-dispersed spectra. It provides automatic data reduction for spectrographs without a bespoke package. Preparing the reduction for a new instrument can be achieved within a working day, significantly less if the pixel to wavelength relation is provided. Nevertheless, the high flexibility of the package comes with the price of less precision, usually dominated by line identification and the corresponding

wavelength solution. Testing the package on HARPS data from a single night indicates radial velocities with a scatter of $3 \, \mathrm{m \, s^{-1}}$ can be obtained without any fine tuning.

Future versions will include more sophisticated procedures for wavelength calibration and cosmic ray removal to give improved functionality and resilience, e.g., the deepCR package (Zhang & Bloom 2019) and xwavecal Brandt et al. (2019). We also plan to expand the compatibility to further RV packages, some given in the Introduction. Further optimization in computational efficiency are possible, e.g., in the extraction process and in parallelization. Due to the extra steps required for the flexibility, extraction needs up to 50% more computation time than the other tested packages for this work.

5. Data access

MRES data used in this publication can be obtained from https://uhra.herts.ac.uk/ under a Creative Commons Attribution license. HARPS data is available from the ESO achive.

ORCID iDs

Ronny Errmann ⓘ https://orcid.org/0000-0002-3683-1952
Neil Cook ⓘ https://orcid.org/0000-0003-4166-4121
David Mkrtichian ⓘ https://orcid.org/0000-0001-5094-3910
William Martin ⓘ https://orcid.org/0000-0003-0757-6689
Fabo Feng ⓘ https://orcid.org/0000-0001-6039-0555
Hugh R. A. Jones ⓘ https://orcid.org/0000-0003-0433-3665

References

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Natur, 536, 437
Anglada-Escudé, G., & Paul Butler, R. 2012, ApJS, 200, 15
Barnes, J. R., Jenkins, J. S., Jones, H. R. A., et al. 2014, MNRAS, 439, 3094
Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Luger, R. 2019, AJ, 158, 164
Brahm, R., Jordán, A., & Espinoza, N. 2017, PASP, 129, 034002
Brandt, G. M., Brandt, T. D., & McCully, C. 2019, arXiv:1910.08079
Buisset, C., Poshyachinda, S., Soonthornthum, B., et al. 2018, Proc. SPIE, 10714, 1071407
Caballero, J. A., Guàrdia, J., López del Fresno, M., et al. 2016, Proc. SPIE, 9910, 99100E
Di Marcantonio, P., Sosnowska, D., Cupani, G., et al. 2018, Proc. SPIE, 10704, 107040F
Feng, F., Lisogorskyi, M., Jones, H. R. A., et al. 2019, ApJS, 244, 39
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Kanodia, S., & Wright, J. 2018, RNAAS, 2, 4
Kaufer, A., Stahl, O., Tubbesing, S., et al. 1999, Msngr, 95, 8
Kramida, A., Ralchenko, Yu., Reader, J. & NIST ASD Team 2018, NIST Atomic Spectra Database (version 5.6.1), https://physics.nist.gov/asd
Kunder, A., Kordopatis, G., & Steinmetz, M. 2017, AJ, 153, 75
Lhospice, E., Buisset, C., Jones, H. R. A., et al. 2019, Proc. SPIE, 11117, 111170Z
Lovis, C., & Pepe, F. 2007, A&A, 468, 1115
Lovis, C., Snellen, I., Mouillet, D., et al. 2017, A&A, 599, A16
Mayor, M., Pepe, F., Queloz, D., et al. 2003, Msngr, 114, 20
Rajpaul, V. M., Aigrain, S., & Buchhave, L. A. 2020, MNRAS, 492, 3960
Ryde, N., Hartman, H., Oliva, E., et al. 2019, A&A, 631, L3
Sosnowska, D., Lovis, C., Figueira, P., et al. 2015, in ASP Conf. Ser. 495, Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), ed. A. R. Taylor & E. Rosolowsky (San Francisco: ASP), 285
Tanvir, T. S. 2018, Finding Close Encounters: Toward a Consensus on the Influence of the Stellar Flybys on the Solar System Over 10 Million Years, https://uhra.herts.ac.uk/handle/2299/20958
Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, Proc. SPIE, 2198, 362
Zechmeister, M., Anglada-Escudé, G., & Reiners, A. 2014, A&A, 561, A59
Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, A&A, 609, A12
Zhang, K., & Bloom, J. 2019, JOSS, 4, 1651