



<b>Publication Year</b>	2016
<b>Acceptance in OA@INAF</b>	2020-05-18T15:37:36Z
<b>Title</b>	Integrated data analysis in the age of precision spectroscopy: the ESPRESSO case
<b>Authors</b>	CUPANI, Guido; D'ODORICO, Valentina; CRISTIANI, Stefano; González-Hernández, Jonay I.; Lovis, Christophe; et al.
<b>DOI</b>	10.1117/12.2231379
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/24930">http://hdl.handle.net/20.500.12386/24930</a>
<b>Series</b>	PROCEEDINGS OF SPIE
<b>Number</b>	9913

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

## Integrated data analysis in the age of precision spectroscopy: the ESPRESSO case

Cupani, Guido, D'Odorico, Valentina, Cristiani, Stefano, González-Hernández, Jonay, Lovis, Christophe, et al.

Guido Cupani, Valentina D'Odorico, Stefano Cristiani, Jonay I. González-Hernández, Christophe Lovis, Sérgio Sousa, Giorgio Calderone, Roberto Cirami, Paolo Di Marcantonio, Denis Mégevand, "Integrated data analysis in the age of precision spectroscopy: the ESPRESSO case," Proc. SPIE 9913, Software and Cyberinfrastructure for Astronomy IV, 99131T (26 July 2016); doi: 10.1117/12.2231379

**SPIE.**

Event: SPIE Astronomical Telescopes + Instrumentation, 2016, Edinburgh, United Kingdom

# Integrated data analysis in the age of precision spectroscopy: the ESPRESSO case

Guido Cupani<sup>a</sup>, Valentina D’Odorico<sup>a</sup>, Stefano Cristiani<sup>a</sup>, Jonay I. González-Hernández<sup>b</sup>,  
Christophe Lovis<sup>c</sup>, Sérgio Sousa<sup>d</sup>, Giorgio Calderone<sup>a</sup>, Roberto Cirami<sup>a</sup>, Paolo Di  
Marcantonio<sup>a</sup>, and Denis Mégevand<sup>c</sup>

<sup>a</sup>INAF – Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy

<sup>b</sup>Instituto de Astrofísica de Canarias, Vía Láctea, 32805 La Laguna, Tenerife, Spain

<sup>c</sup>Université de Genève, 51 Chemin des Maillettes, 1290 Versoix, Switzerland

<sup>d</sup>Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das  
Estrelas, PT4150-762 Porto, Portugal

## ABSTRACT

The Echelle SPectrograph for Rocky Exoplanets and Stable Spectral Observations (ESPRESSO) is an ultra-stable spectrograph for the coudé-combined focus of the VLT. With its unprecedented capabilities (resolution up to  $\sim 200,000$ , wavelength range from 380 to 780 nm; centimeter-per-second precision in wavelength calibration), ESPRESSO is a prime example of the now spreading science machine concept: a fully-integrated system carefully designed to perform direct scientific measurements on the data, in a matter of minutes from the execution of the observations. This approach is motivated by the very specific science cases of the instrument (search for terrestrial exoplanets with the radial velocity method; measure of the variation of fundamental constants using the spectral signatures of the inter-galactic medium) and is achieved by a dedicated tool for spectral analysis, the data analysis software or DAS, targeted to both stellar and quasar spectra. In this paper, we describe characteristics and performances of the DAS, with particular emphasis on the novel algorithms for stellar and quasar analysis (continuum fitting and interpretation of the absorption features).

**Keywords:** data analysis, high-resolution spectroscopy, ESPRESSO, ESO VLT, instrument pipeline, science machine

## 1. INTRODUCTION: ESPRESSO IN A NUTSHELL

Several science cases in optical spectroscopy are pushing the limits of the existing instruments, depending on extreme sensitivity, high resolution and exceptional thermo-mechanical stability. One of the instruments designed to satisfy these requirements is the Echelle SPectrograph for Rocky Exoplanets and Stable Spectral Observations (ESPRESSO),<sup>1</sup> designed for the ESO VLT of Cerro Paranal, Chile. ESPRESSO is targeted to a set of very specific science cases:

- The detection of extrasolar planets from observed fluctuations in the radial velocity of stars;
- The measure of a possible variation in the value of dimensionless physical constants, namely the fine structure constant  $\alpha$  and the proton-to-electron mass ratio  $\mu$ ;
- The study of the chemical composition of stars in local galaxies;
- The study of the metal-poor stars in our own Galaxy;
- The asteroseismic study of stellar oscillation, correlated with stellar variability;
- The investigation of the physical and chemical properties of the intergalactic medium (IGM) and their connection with galaxy evolution.

---

Send correspondence to Guido Cupani – E-mail: cupani@oats.inaf.it, Telephone: +39 (0)40 3199 235.

Software and Cyberinfrastructure for Astronomy IV, edited by Gianluca Chiozzi,  
Juan C. Guzman, Proc. of SPIE Vol. 9913, 99131T · © 2016 SPIE  
CCC code: 0277-786X/16/\$18 · doi: 10.1117/12.2231379

Table 1. ESPRESSO technical specification for the different observational modes (singleHR: 1 UT, high resolution; multiMR: 4 UTs, medium resolution; singleUHR: 1 UT, ultra-high resolution).<sup>1</sup>

Parameter	singleHR	multiMR	singleUHR
Wavelength range (nm)	380-780	380-780	380-780
Resolving power	134,000	59,000	225,000
Aperture on sky (arcsec)	1.0	$4 \times 1.0$	0.5
Average spectral sampling (pixels)	4.5	5.5 (binned $\times 2$ )	2.5
Spatial sampling (pixels)	9.0 (4.5)	5.5 (binned $\times 4$ )	5.0
Total efficiency	0.11	0.11	0.05
Instrumental wavelength precision	$<10 \text{ cm s}^{-1}$	$\sim 1 \text{ m s}^{-1}$	$<10 \text{ cm s}^{-1}$

All these science cases rely on the acquisition of high-resolution spectra of stars or quasars, the latter used as background sources to pierce the IGM through a wide range of redshifts. The first two science cases, in particular, require an unprecedented precision and accuracy in wavelength calibration to improve on the results obtained with the existing instruments. Concerning the search for extrasolar planets, the current detection limit set by the HARPS<sup>2</sup> and HARPS-N<sup>3</sup> spectrographs stands at the order of  $1 \text{ m s}^{-1}$  and falls short of one order of magnitude of the  $10 \text{ cm s}^{-1}$  limit required to detect Earth-like planets in the habitable zone. Similarly, the upper limit in both  $\alpha$  and  $\mu$  variation is of the order of some parts per million<sup>4-6</sup> and can be further reduced only by reaching a  $<1 \text{ m s}^{-1}$  precision in wavelength calibration. This improvement is also pivotal to the next generation of spectrographs for extremely large telescopes, like E-ELT HIRES,<sup>7</sup> which are going to address even more delicate science cases (such as the direct measurement of the Universe accelerated expansion, the so-called Sandage test<sup>8,9</sup>) and will require a wavelength precision of the order of  $2 \text{ cm s}^{-1}$ .

ESPRESSO is a fiber-fed echelle spectrograph to be installed in the combined coudé laboratory of the VLT, placed underground on Cerro Paranal. It will receive light from the four Unit Telescopes (UT) through optical coudé trains, interfaced to four dedicated front-end units, and will be able to operate at different resolutions with either a single UT or up to four UTs at once. Its optical design adopts several solutions to combine high resolution and high sensitivity, including fiber scrambling, anamorphic pupil slicing, and a two-arm layout with separate blue and red detectors. The optical bench is lodged within a vacuum vessel to stabilize pressure with a precision of  $1 \mu\text{Pa}$  during the night; three nested thermal enclosures will keep temperature fluctuations within  $1 \text{ mK}$  during the same time scale. A laser frequency comb (LFC), consisting in a pattern of narrow, equally spaced emission lines generated by a mode locked laser, will provide a simultaneous reference for wavelength calibration through a dedicated fiber. Other technical specifications for the instrument are given in Tab. 1. The development has now reached the integration phase; the first light is foreseen for the first half of 2017.

In this paper we detail how the data produced by ESPRESSO will be handled after their reduction by a dedicated software tool for data analysis. We describe both the architecture of the software (Sec. 2) and the algorithms which have been implemented to manage the requirements posed by the instrument science (Sec. 3). These algorithms either builds on existing techniques or introducing brand-new methods, as in the case of continuum fitting for quasar spectra, and their relevance extends beyond the immediate scope of the instrument (Sec. 4).

## 2. INTEGRATED DATA FLOW SYSTEM

The need for a dedicated tool for data treatment was underlined since the first stages of the ESPRESSO design phase. This tool was meant to complement the standard tool for data reduction and to enforce a new paradigm, named the *science machine* approach: the idea that an instrument should be fed by the end user with setup parameters based on the scientific results to be achieved, and should produce usable, science-grade data with minimal external supervision. Ideally, the end user will be able not only to reduce the data, but also to measure relevant quantities (e.g. radial velocities, line equivalent widths and Voigt parameters, etc.) within minutes from the execution of the observations. This novel approach is favored by the specific science cases of ESPRESSO, and is likely to become more widespread as astronomical spectroscopy evolves into a precision science.

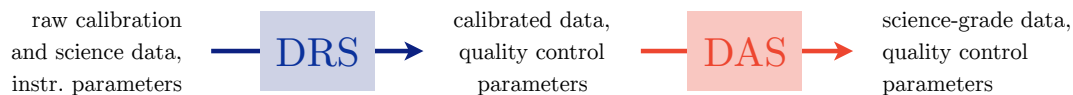


Figure 1. Scheme of the ESPRESSO data processing flow, from the raw products of the observations to the science-grade products.

## 2.1 Data Analysis within the ESO environment

The ESPRESSO Data Analysis Software (DAS) is a collection of ANSI-C plugins (*recipes*) mostly implemented with the ESO Common Pipeline Library (CPL).<sup>10</sup> Each recipe performs a distinct and complete operation on the reduced data; they all can be run as stand-alone modules or in cascade (*pipeline*) using one of the interfaces provided by ESO (EsoRex, a command-line interface; Gasgano, a graphical user interface with browser capabilities; and Reflex, a workflow-management GUI: see also Sec. 2.2). The recipe are complemented by a collection of Python scripts to display the results of the operations. The DAS will be distributed under the GNU General Public License after the instrument commissioning.

The DAS is seamlessly integrated within the ESO Data Flow Systems (DFS), which handles in a continuous loop the flow of data across the ESO instruments, the ESO archive, and the end users. Conceptually, the ESO data flow includes four steps:

1. The preparation and submission of observation proposals, which are evaluated and approved by the Observing Programmes Committee;
2. The preparation of the observing blocks through a dedicated interactive software;
3. The execution of the observing blocks through the observation preparation software, and the subsequent population of the archive with raw calibration and science data;
4. The processing of raw data (reduction and analysis), which actively contributes to monitor the instrument health status and the quality of the observation output, and generates science-grade data products also suitable to be stored in the archive.

While the data reduction is routinely performed by instrument pipelines, validated and maintained by ESO (e.g. for the UV and optical echelle spectroscopy, the UVES pipeline<sup>11,12</sup> and the X-shooter pipeline<sup>13</sup>), until ESPRESSO the data analysis was delegated to general-purpose software suites (e.g. ESO-MIDAS,<sup>14</sup> whose support has been discontinued since 2012). The ESPRESSO DAS is the first dedicated data analysis software for an ESO instrument, meant to work in close interaction with the ESPRESSO Data Reduction Software (DRS) according to the scheme in Fig. 1.<sup>15–18</sup> Both tools will be available at the VLT control room in Paranal, Chile (the DRS will be automatically triggered by the injection of new observational raw data into the archive, while the DAS will be at the visitor astronomers' disposal for offline processing); they will be also provided as downloadable self-installing kit for the Fedora Linux and Mac OS X environments.

## 2.2 Structure and operation of the ESPRESSO DAS

The complete set of DAS recipes is shown in Fig. 2. Recipes are divided into two main *branches*, depending on them being aimed to stellar or quasar spectra; some general-use recipes are shared between the two branches. Recipes can be arranged in several ways to handle the different requirements of stellar and quasar analysis; for the typical use case, four pipelines have been outlined, as shown in Fig. 2.

Data analysis is generally an iterative process, in which results are refined on several cycles through the same procedures. A typical example is the analysis of the absorption lines in a quasar spectrum, which involves a number of tightly-interlocked operations and cannot be easily schematized as a linear procedure (see Sec. 3.4). To comply with such peculiarity, the DAS is designed to be highly modular and customizable. In particular:

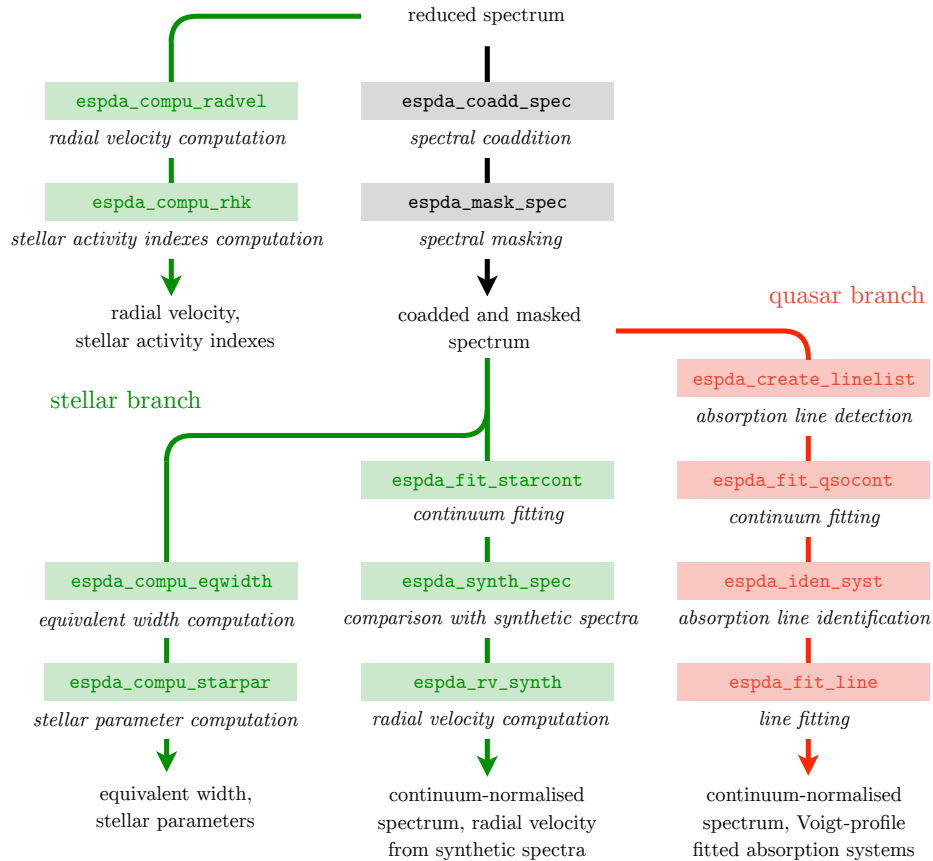


Figure 2. Conceptual plan of the DAS. Recipes are depicted as shaded blocks (green: stellar branch recipes; red: quasar branch recipes; gray: shared recipes), with captions describing the operation performed by each recipe. The four pipelines implemented as Reflex workflows are depicted with arrows (three for the stellar branch, one for the quasar branch).

- At the recipe level, each relevant sub-operation is delegated to an independent ANSI-C function, to make tasks equally accessible by different recipes;
- At the pipeline level, the arrangement is designed to be fluid: in all applicable cases, recipes are meant to be interchangeable and to take input not only from the upstream recipes, but from the downstream ones as well.

While the internal structure of recipes is fixed by algorithmic constraints, the pipeline structure (though optimized for the ESPRESSO science cases discussed in Sec. 1) is susceptible of adjustment by the end users, to accommodate specific needs.

The preferred interface to the DAS recipes is the ESO Reflex environment,<sup>19</sup> developed within the Kepler workflow engine. The four pipelines depicted in Fig. 2 are all equipped with a dedicated Reflex workflow; a snapshot of the main canvas for the quasar branch pipeline is shown in Fig. 3. The interaction with the pipelines is fully managed by the Reflex GUI: recipes are executed in sequence and their products are consistently propagated across each other. The overall process is controlled by the play/pause/stop buttons on its toolbar. A typical analysis session goes as follows:

- Input data are retrieved from the hard drive and automatically organized into data sets (typically, one for each scientific target, including all the reduced spectra and the required reference data). The user can inspect the data sets and select them for analysis.

- Data sets are processed separately. Information is passed across the recipes as lists of files, which specify the location of the required data and their nature (e.g. spectrum, list of absorption line, etc.), and lists of parameters, which specify the recipe setup. Each recipe takes a list of files in input and gives a list of files in output. Recipe parameters may be initialized to particular values directly on the Reflex canvas or from the recipe configuration window.
- When convenient, the products of a given recipe are displayed in a pop-up window. Such windows are generated by Python scripts and are fully interactive, allowing for data inspection and parameter adjustment. When the window appears, processing is paused: the user may decide to re-execute the recipe with a different parameter setup, or to continue the execution along the pipeline.
- When science-grade data have been produced (typically at the end of the pipeline, but also after intermediate steps) they are stored in the hard drive. The user may inspect them and check their consistency.

An example of Reflex pop-up window is shown also in Fig. 3. In this specific case, the user can zoom into the plot (e.g. to check the details of the processed spectrum), select points on the plot area with mouse clicks (e.g. to pinpoint absorption lines), and manually adjust the parameters in the text boxes. Other pop-up windows are able to display editable tables of absorption lines with guess or best-fitting Voigt parameters. Python events on pop-up windows, such as mouse clicks and table cell editing, are converted into recipe setup parameters, and may trigger the re-execution of the recipes.

In addition to science-grade data products, Reflex saves all the required ancillary data (including lists of files and parameter, configuration files and log files) in a dynamically updated directory tree. With proper scripting, this information may be used to mimic Reflex operations via a command-line interface, using EsoRex and a Python interpreter. It is thus possible to operate the DAS pipelines at different depth levels, without sacrificing control for the sake of automation.

### 3. NOVEL SOLUTIONS FOR SPECTRAL ANALYSIS

The general idea behind the DAS is to unify within well-tested algorithms the operations that the end users normally perform by eye or with custom-made, minimally-validated software routines. This is done in two ways:

1. By consolidating existing software, such as the ARES code<sup>21,22</sup> to compute the equivalent width of lines (Sec. 3.2.1), and validating it to work on ESPRESSO data;
2. By developing new code building on long-time experience with manual analysis: it is the case of the newly-formalized technique to determine the continuum emission in quasar spectra (Sec. 3.4.1).

We give in this section a detailed description of the most relevant algorithms for both the stellar branch and the quasar branch pipelines.

#### 3.1 Conservative approach to spectral coaddition

We outlined the problem of coadding multiple spectra of the same scientific target in a previous SPIE Proceedings article.<sup>16</sup> Coaddition is typically performed by the end user on the rebinned spectra issued from data reduction, whose spectral orders have already been merged and optimally extracted into a 1D array with a fixed wavelength grid. Exposures of the same target taken at different epochs are generally rebinned into different wavelength grids. Even assuming that the spectral format of the instrument is perfectly stable, the wavelengths associated to each pixel of the detector change with time due to the motion of the Earth with respect to the barycentric coordinate frame of the Solar System. For this reason, different exposures cannot be coadded pixel by pixel but require a further rebinning to be combined into a single spectrum. The problem with rebinning is that it introduces a spurious correlation in the noise values across adjacent bins, with the effect of underestimating the error on the flux measurement. This seriously undermines the possibility to use error-weighted statistical tools, like the  $\chi^2$  test, to estimate the goodness of fit on spectral features (see Sec. 3.4).<sup>23</sup> It can be shown that also when the wavelength grids of the different exposures are coincident, a second rebinning after the one already



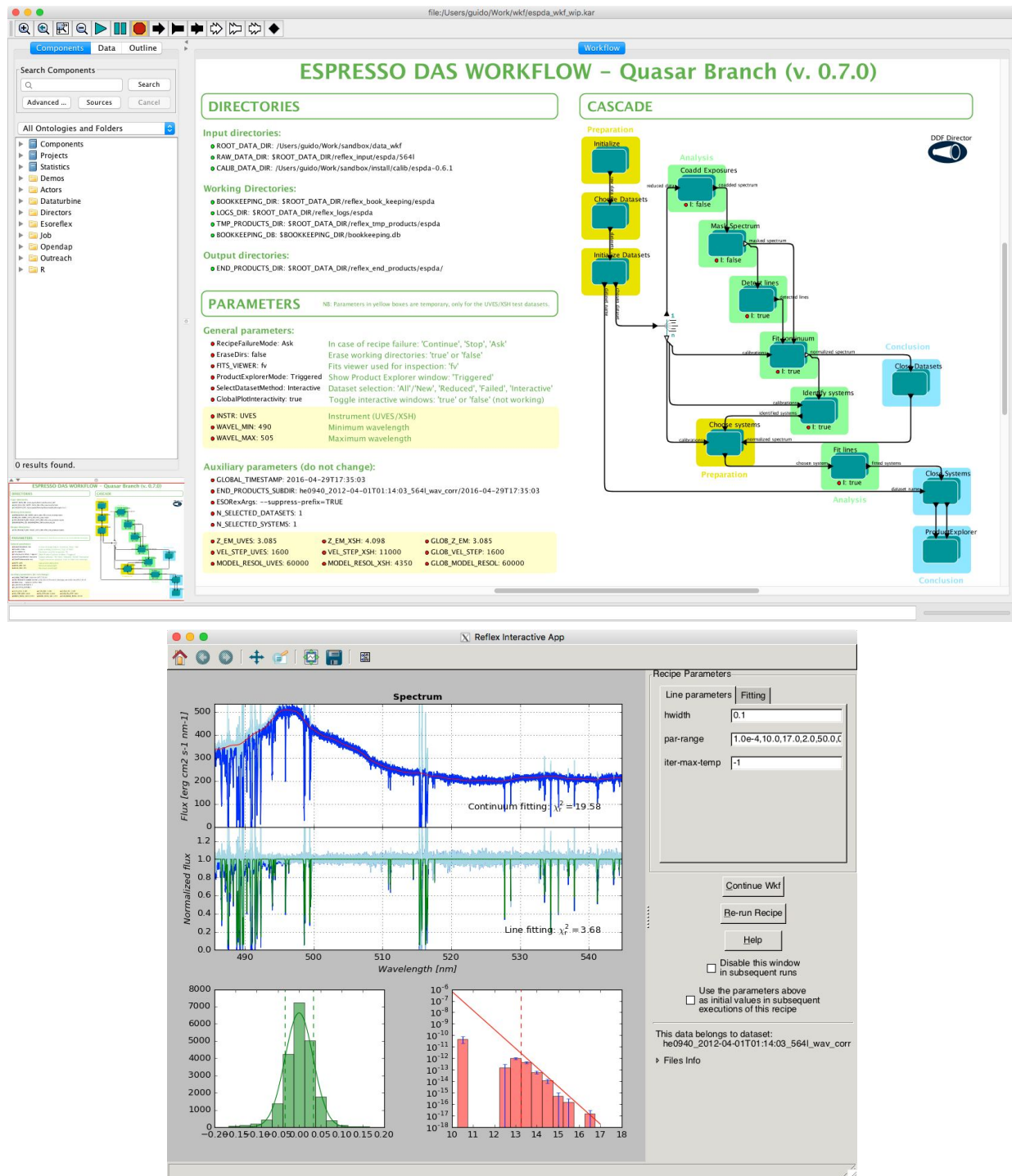


Figure 3. The Reflex GUI. *Top image:* canvas for the quasar branch pipeline. On the left side, the main operational parameter are listed (for data retrieval and recipe setup). The pipeline is shown on the right side: each green-highlighted box corresponds to a recipe and contains the Reflex actors which execute it and display its products. *Bottom image:* Reflex pop-up window for recipe `espda_fit_qsocont`. The panels on the left side show (counter-clockwise from top) the estimated continuum emission, the continuum-normalized spectrum (with fitted absorption lines), the residual distribution of the line/continuum fitting procedure (see Sec. 3.4) and the distribution of column densities from Voigt-profile line fitting (compared with the theoretical distribution from Ref. 20). Adjustable recipe parameters appear in the right side.



performed by the data reduction generally alters the distribution of errors in the final spectrum (we reserve to discuss this issue in a future article).

Two different approaches are adopted by the DAS to cope with this problem:

- The *coadded spectrum* is computed using the non-rebinned spectra issued from data reduction, in the form of 2D matrices containing each spectral order as a separate 1D array\*. Pixels in the 2D matrices correspond one-to-one to pixels in the detector and retain their individual wavelengths, computed for each exposure from the simultaneous LFC reference. A fixed wavelength grid is adopted for the final spectrum, and the contributions from all exposures are binned into this grid without a prior rebinning of the individual spectra, to minimize the correlation across bins. The final spectrum is formatted as a VOTABLE with WAVE, FLUX, and ERR\_FLUX columns.
- In addition to the coadded spectrum, a *combined spectrum* is also created, by collecting all the pixels of the 2D matrices. This spectrum is also formatted as a VOTABLE with WAVE, FLUX, and ERR\_FLUX columns, with an additional PIXSIZE column to retain the wavelength extension of the original pixels. Since this table is produced without any rebinning, it can be safely used as a reference to test the goodness of fit in the subsequent analysis. Within this second approach, the information content of the multiple exposures accounts for an effective oversampling of the spectrum: since the individual 2D matrices are defined on independent wavelength grids, their combination results in a much finer grid for the final spectrum, which is also beneficial to the fitting of spectral lines<sup>†</sup>.

Both approaches are implemented in the DAS recipe `espda_coadd_spec`. The algorithm to produce the coadded spectrum is modeled on the “drizzling” technique used to combine HST images.<sup>25</sup> The flux in each bin of the final spectrum is obtained from a weighted average of the contributions of all pixels from the original exposures that (at least partially) overlap the wavelength range of the bin (Fig. 4). For each contribution, the weights are proportional to the overlap and inversely proportional to the square of the errors. The recipe also monitors the signal-to-noise ratio (SNR) and the error statistics of the coadded spectrum, comparing the to the RMS of the contributions, and raises a flag when a significant discrepancy occurs. Optional kappa-sigma clipping of the contributions and rescaling of the flux values are available to amend any remaining defect in the reduced spectra.

## 3.2 Quick computation of stellar parameters

The ESPRESSO DAS will provide a fast and reliable estimation of stellar parameters for solar-type stars. To achieve this DAS contains two dedicated recipes in a specific workflow. One of the recipes (`espda_compu_eqwidth`) compute the line strengths for metal absorption lines using the equivalent width (EW) method. The second recipe (`espda_compu_starp`) uses well proven line-ratio calibrations to estimate precise effective temperatures and use a similar iron line calibrations to obtain the iron abundance which is typically used as a proxy for the stellar metallicity.

### 3.2.1 EW determination

The recipe `espda_compu_eqwidth` is a compilation of many of the functions used in the ARES code,<sup>21,22</sup> which were rewritten using the CPL libraries and following ESO guidelines for its implementation. Essentially, the recipe tries to replicate and automatize the normal approach to absorption line strengths measure based on interactive routines. The main steps are:

1. Determination of the local continuum;
2. Identification of the position and depths of the absorption lines in the local spectrum;

---

\*Thanks to ESPRESSO’s convenient optical layout and fiber scrambling, the image of the entrance slit on the detector is not tilted along orders, and it retains no spatial information. As a consequence, the individual orders can be optimally extracted by summing photon counts along pixel columns.

<sup>†</sup>A similar effective oversampling is routinely adopted to model the sky emission lines in the near-infrared bands.<sup>24</sup>

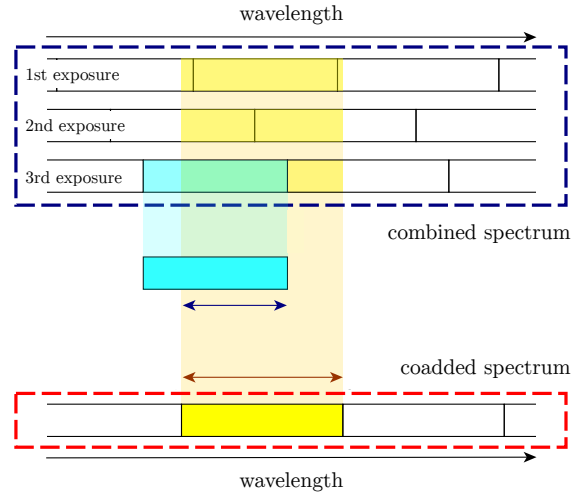


Figure 4. Scheme of the “drizzling”-like algorithm to produce the coadded spectrum. Three non-rebinned spectra issued from data reduction (blue dashed square) are rebinned into a fixed-wavelength grid spectrum (red dashed square). Each bin of the coadded spectrum (in yellow) takes contributions from many different pixels of the original spectra (yellow-shaded region). In the particular case of the cyan pixel, the ratio between the blue arrow and the red arrow defines the overlap used to weight the contribution.

3. Non-linear least-square fitting of the identified lines with Gaussian profiles;
4. Determination of the EW and error estimation based on the numerical fit.

The recipe works on a spectral chunk of a given wavelength range, centered on each individual line to be measured. The first step for each line is to perform a local continuum determination by iteratively fitting a 2<sup>nd</sup> order polynomial function to the local spectrum. At each iteration, points below the fit are removed in order to keep the best points for the continuum. A rejection parameter, strongly dependent on the SNR of the spectrum, controls the selection of points; by default, all values below  $1 - 1/\text{SNR}$  are removed.

The algorithm for the automatic identification of absorption lines is an important step for the correct computation of the EW. As in ARES, this process is based on the use of derivatives of the local spectrum, since an absorption line will generally correspond to local minima in the local spectrum. However, there are cases when absorption lines are severely blended and the correct identification cannot be done just looking at the minima. For this reason, the algorithm also uses the second and third derivative of the local spectrum, which will correspond to inflection points. These points are successfully used to identify the lines even in cases of extremely blended features (see Fig. 3 of Ref. 21). As the noise strongly propagates through the numerical derivatives, it should be properly smoothed out. The noise smoothing length generally depends both on the SNR of the spectrum (the lower the SNR, the higher the value should be) and on the spectral resolution (or pixel scale). Since the ESPRESSO resolution is fixed by its observational mode, we only consider the dependence with SNR.

Gaussian-profile fitting is performed automatically on multiple lines, using the `gsl_multifit` package from the GNU Scientific Library (GSL).<sup>26</sup> The actual measurement of EW is done from the best-fit Gaussian parameters. It should be noted that the recipe was developed to compute the EW of weak lines, and it is optimized to work in combination with the `espda_compu_starpar` (see Sec. 3.2.2).

### 3.2.2 Effective Temperature and [Fe/H] for solar-type stars

The ESPRESSO DAS will allow the user to estimate effective temperature and iron abundance on the fly once an ESPRESSO spectrum is observed and reduced. These estimations computed by the recipe `espda_compu_starpar` are based on well proven calibrations which work very well for a restricted, but still significantly wide, parameter space.

The algorithm for the effective temperature estimation relies on the use of a line-ratio technique.<sup>27</sup> Although the typical technique relies on line-depth ratio, in this case the recipe uses line EW ratio which in principle retains more information about the strength of an individual line. The calibration used in this recipe was initially derived in Ref. 28 which made use of a unique calibration sample composed of 451 stars with high quality spectra observed with HARPS. The quality of this calibration does not only rely on the quality of the spectra but also importantly on the homogeneously derived spectroscopic parameters.<sup>21</sup> These together allowed to calibrate 432 line-ratios with individual precisions better than 120 K with the best ratios having precision of  $\sim 50$  K.

The determination of the iron abundance, which is typically used as a proxy for the stellar metallicity, is also derived by this recipe following a similar approach. Here the recipe also relies on a calibration, performed in this case on specific, well-selected iron absorption lines. The calibration process and the selection of these lines are described in detail in Ref. 29; essentially, they are a compilation of Fe I lines: Fe II lines were neglected as they are strongly dependent on the stellar surface gravity for solar type stars. The calibration was performed only for lines weaker than 75 mÅ to avoid dependence with the microturbulence which becomes important for relatively stronger lines. In this way, the remaining Fe I lines have only large-order dependencies on the iron abundance and the effective temperature. The latter is independently derived by the line-ratio technique described before. As for the line-ratio calibrations this metallicity calibration was derived using the same very high quality sample which allowed a calibration of 151 Fe I lines; several tests on independent samples shows very consistent results.<sup>29</sup>

Since the recipe relies on these calibrations, the user should be aware of the strict parameter space of validity for which the calibration were obtained. In this case the original calibrations are suitable to estimate parameters for dwarf FKG stars with effective temperatures ranging from 4500 K to 6500 K and with [Fe/H] ranging from  $-0.8$  dex to  $0.4$  dex. If the user tries to apply the procedure to other types of star, the output will be provided without any real validation. However, the errors for these cases are expected to be significantly larger than the ones that we obtain for stars within the parameter space validity and therefore can be used as a reference to check the validity of the estimations. Although the current original calibrations are valid for dwarf stars, new calibrations are being prepared to include also evolved stars. These new calibrations will have a wider parameter space and will be valid for FGK-dwarfs and GK-giant stars in the following ranges:  $4500 \text{ K} < T_{\text{eff}} < 6500 \text{ K}$ ,  $2.5 < \log g < 4.9$ ,  $-0.8 < [\text{Fe}/\text{H}] < 0.5$ .

### 3.3 Continuum fitting in stellar spectra and comparison with synthetic spectra

Normalizing echelle spectra after rebinning and merging is a challenging task due to the low flux counts at the edges of each spectral order. In order to overcome this difficulty the DAS will be equipped with an algorithm (included in the recipe `espd_fit_starcont`) to normalize the continuum of stellar echelle spectra order by order. The recipe selects a spectral order from the combined spectrum (see Sec. 3.1) and fits the stellar continuum with a  $N$ -order polynomial, using robust linear regression as implemented in the `gsl_multifit` package from the GSL.<sup>26</sup> The fitting procedure is done in several iterations after rejecting those points above and below a high and low threshold (a few times the standard deviation of the non-rejected flux points). This recipe provides two output products:

- A combined spectrum including the continuum fit, the final selection of points of the fitting procedure and flux errors of the fitting stellar continuum fluxes;
- A rebinned spectrum, obtained from the combined one with the procedure described in Sec. 3.1.

Fig. 5 shows an example of the fitting of order spectra of stars of different spectral types.

The technique to derive chemical composition of stars is based on the comparison of the equivalent width of individual absorption lines measured in the observed spectrum with synthetic measurements using the curve-of-growth method (an introduction to this method can be found in Ref. 30). Alternatively, the comparison of the whole spectral region with spectral synthesis is another approach but involves a calculation that includes all the spectral lines in the region of interest. The comparison with single synthetic spectra could provide an estimate of the metallicity of an observed star. The DAS pipeline includes a routine aiming to only compare the observed spectrum with one single interpolated synthetic spectrum. In order to perform the comparison, we first interpolate

the synthetic spectrum from a grid of synthetic spectra, using a given set of stellar and broadening parameters. The grid of synthetic spectra is computed with Kurucz model atmospheres.<sup>31</sup> This interpolation procedure is performed by recipe `espsda_synth_spec`, using a selected set of stellar parameters. This spectrum is convolved with three broadening profiles for stellar rotation, macroturbulent velocity and instrumental broadening, and then rebinned to the wavelength array of the observed normalized merged spectrum. The output product of this recipe is the rebinned synthetic spectrum. The observed spectrum needs to be corrected for radial velocity with another recipe, `espsda_rv_synth`, which performs a cross-correlation of the observed spectrum with the synthetic spectrum. The radial velocity is determined with a gaussian fit to the cross-correlation function (CCF). Both the CCF and the observed stellar spectrum corrected for radial velocity are given as output products of this recipe.

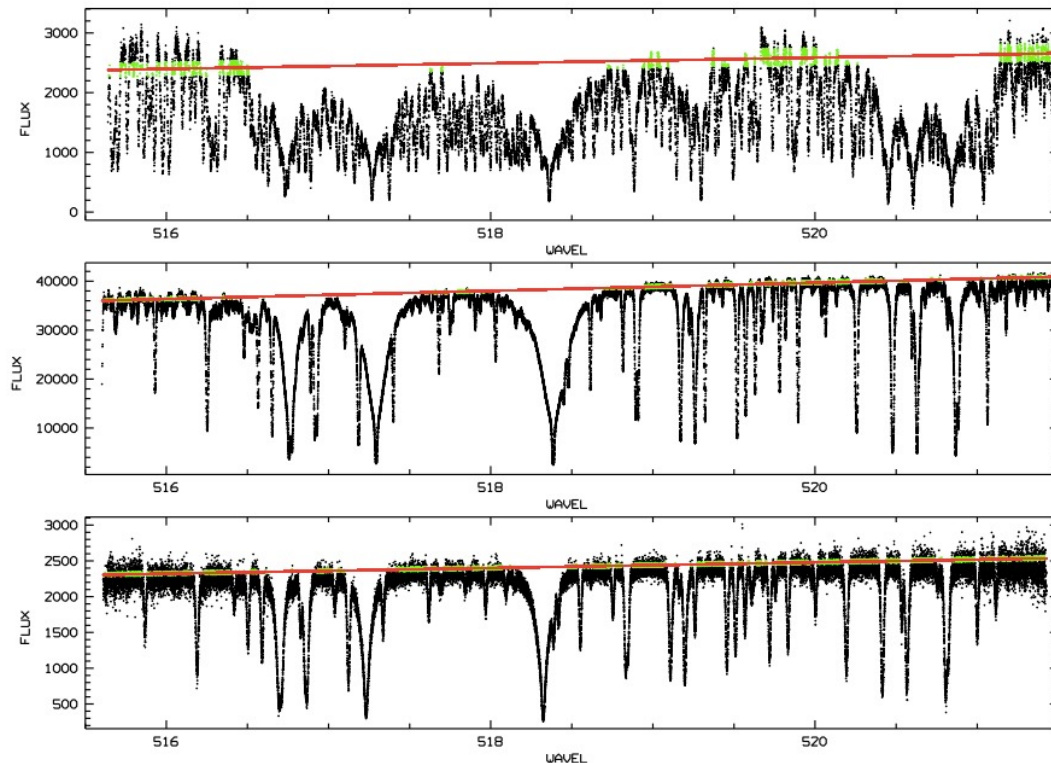


Figure 5. Example of the result of the recipe `espsda_fit_starcont` using HARPS spectra of K7V (top panel), G6V (middle panel) and F7V (bottom panel) stars. We show the continuum fitting of order 44 (spectral region of MgIb triplet) of ten individual spectra of these stars coadded using the recipe `espsda_coadd_spec`. Green points show the selected continuum points after several iterations, and red line is a fit of order 1.

### 3.4 Iterative continuum and absorption line fitting in quasar spectra

The interpretation of the absorption features imprinted on a quasar spectrum by the IGM along the line of sight is an intrinsically challenging task, which involves at least three distinct operations:

- determining the unabsorbed quasar emission spectrum, the so-called *continuum*;
- detecting and identifying the absorption lines by finding the ionic transitions that produced them;
- fitting a Voigt profile to the identified lines, from which the physical properties of the IGM can be inferred.

Each operation ideally requires previous knowledge obtained by the other two: the continuum is best determined by iteratively fitting and removing all the absorption lines; the lines can be safely identified only when their

parameters have been constrained by continuum and Voigt-profile fitting; and the Voigt-profile fitting itself requires information about the line wavelengths and the continuum level to be performed. An iterative approach is needed to handle the problem consistently.

It is common practice to firstly determine the continuum emission of the quasar by masking or extrapolating across absorption lines, and subsequently model the lines on the continuum-normalized spectra. The effectiveness of the procedures strongly depends on the possibility to isolate absorption lines and to safely assess as continuum the emission component between them. This is not always the case: in fact, a significant portion of the visible spectrum of medium- to high-redshift quasar is dominated by the *Lyman- $\alpha$  forest*, a complex pattern of frequently blended, occasionally saturated absorption lines arising from absorption by the neutral hydrogen in the IGM, bluewards of the Lyman- $\alpha$  emission line of the quasar itself. Continuum determination in the Lyman- $\alpha$  forest has always been tentative, and is routinely performed by eye (selecting regions arguably free from absorption and fitting a smoothing functions across them) or making assumptions on the expected quasar emission (e.g. through a principal component analysis over a catalogue of previously computed continua<sup>32</sup>). Iterative techniques based on simultaneous line-and-continuum fitting have been in fact developed, but only to work on limited regions of the Lyman- $\alpha$  forest (e.g. around damped Lyman- $\alpha$  systems<sup>33</sup>).

The ESPRESSO DAS extends the iterative approach to the entire spectrum. Continuum fitting, line identification, and line fitting are still treated as individual operations (performed by recipes `espd_fit_qsocont`, `espd_iden_line`, and `espd_fit_line`, respectively: see Fig. 2), but they significantly overlap and can be executed in whichever order, reinforcing each other. In its basic form (implemented in the DAS quasar branch pipeline), the procedure starts from the continuum determination, which is obtained through a first, minimal-guess fitting of the absorption lines, and develops through the line identification and a refined line fitting, which corrects the initial guess and provides information to refine the continuum as well. Each steps converges to its solution by iterative means, and the three steps together can be globally iterated by executing the pipeline multiple times using previous products as input.

### 3.4.1 Continuum fitting

The continuum fitting algorithm splits the spectrum into a large number of chunks and determines the unabsorbed emission level in every one of them as a single value for the entire chunk<sup>‡</sup>; the final continuum is then determined by computing a cubic B-spline across these points. In each chunk, the continuum level is recovered by iteratively fitting all the detected absorption lines. The generic iteration  $i$  goes as follows:

1. A guess continuum value is extracted from the statistics of the flux spectrum within the chunk (typically, taking the median of the flux). Additional assumptions about the global continuum shape (e.g. a combination of power laws<sup>34</sup>) can be used in the first iteration.
2. Absorption lines are fitted with Voigt profiles after normalizing the flux to the guess continuum. Only the strongest  $i$  lines are fitted, until  $i$  equals the number of lines detected in the chunk.
3. After fitting the lines, the unabsorbed flux spectrum is approximately reconstructed dividing the observed flux spectrum by the results of the line fit. The result is used in the next iteration to update the guess continuum estimation.

The algorithm is said to have converged when all lines have been fitted and the resulting unabsorbed flux spectrum is consistent,  $\chi^2$ -wise, with the B-spline fitted final continuum. Within the Lyman- $\alpha$  forest, it is not always possible to explain absorption in terms of detected lines, and ad hoc assumptions are required. A significant bias in the continuum placement during the iterations may in fact lead to an uncontrolled drift away from convergence, resulting either in underestimation (when the amount of absorption compromises the reliability of the flux median as a proxy for the continuum) or overestimation (when a defective line fitting produces large positive residuals in the estimated unabsorbed spectrum). An effective solution to this problem is to enhance the flux spectrum by a redshift-dependent optical depth term, which models the residual Lyman- $\alpha$  absorption not accounted for by the lines that have already been fitted. This effective optical depth is due to both resolved

<sup>‡</sup>This is generally a good assumption if the span of the chunk is of the order of  $10^3 \text{ km s}^{-1}$ .



and unresolved lines and is re-estimated at each iteration from the column density distribution of the fitted lines, decreasing to zero as more lines are fitted. An example of the results is visible on the Reflex window shown in Fig. 3.

### 3.4.2 Line identification

The line identification algorithm is an extension of the technique described in Ref. 35 to detect *absorption systems*, i.e. groups associated lines arising in different hydrogen and metal-line absorbers at the same redshift. It can be outlined as follows:

1. A list of detected lines, with observed wavelengths  $\lambda_l$ , is compared with a first list of known transition, with laboratory wavelengths  $\lambda_t$ . The redshift values  $z_{lt} = \lambda_l/\lambda_t - 1$  are computed for all possible line-transition associations.
2. The sorted list of  $z_{lt}$  is scanned with a running window to find possible internal coincidences. When a coincidence is found, a candidate absorption systems is said to be detected at a redshift  $z_c$ , defined as the average of the coinciding values  $z_{lt}$ .
3. Candidate systems are checked against a set of consistency rules, mostly based on the expected ratio between line strength (as determined during the continuum-fitting procedure). Lines that miss the check are discarded; whole systems can be discarded as well if too many lines miss the check.
4. When a system has been confirmed, the original list of lines is compared with an expanded list of known transitions, with laboratory wavelengths  $\lambda'_t$ . The new redshift values  $z'_{lt} = \lambda_l/\lambda'_t - 1$  are checked to find additional coincidences with  $z_c$  within a narrower window. If found, the consistency rule are applied a second time to the expanded systems.

The line associations issued by such algorithm are suitable to be more accurately fitted with Voigt profiles as legitimate absorption systems.

### 3.4.3 Line fitting

The line fitting algorithm adopts a Levenberg-Marquardt approach to determine the best-fit Voigt parameters of a line or a group of lines. It operates from guess values of the line parameters (typically, the redshift  $z$ , the column density  $N$ , the Doppler broadening  $b$ , and the turbulence broadening  $b_{\text{turb}}$  of each line), which are transformed into proper independent variables for the Voigt function and optimized by minimizing the  $\chi^2$  between the observed line profile and the theoretical one. When called from recipe `espda_fit_qsocont`, the algorithm operates on a relatively crude choice of guess parameters, based on minimal assumptions on the line nature ( $z$  is computed under the hypothesis that the lines are Lyman- $\alpha$ , C IV  $\lambda 1550$ , or Mg II  $\lambda 2803$ , depending on the region of the spectrum;  $N$  is approximated from the line peak flux; the broadening is fixed as  $b = 20 \text{ km s}^{-1}$ ,  $b_{\text{turb}} = 0 \text{ km s}^{-1}$ ). On the contrary, when called from recipe `espda_fit_line`, guess parameters are set taking into account the line identification performed by recipe `espda_iden_syst`, as well as the specific choices of the user (which are set through a dedicated pop-up window in Reflex). In this latter case lines are fit as absorption systems (rather than by unspecified groups) and constraints may be applied between line parameters during the fitting procedure, to effectively reproduce the physical state of the absorbers.

The accuracy of the line fitting algorithm has been tested on simulated lines at different values of line parameters  $N$  and  $b$ .<sup>17</sup> The flux density profile  $I(\lambda)$  of the lines was simulated by interpolating the Voigt function on different wavelength grids, designed to mimic the structure of both non-rebinned and rebinned spectra (see Sec. 3.1 and Fig. 4). Noise was modeled on a pixel-by-pixel basis as a normally-distributed random variable added to the flux density, with variance  $\sigma^2(I) = (I + B)t_{\text{exp}} + \text{RON}$ , where  $B$  models the background noise, RON models the instrument read-out noise, and  $t_{\text{exp}}$  is an artificial exposure time term needed to achieve the desired  $\text{SNR} = \langle I/\sigma^2(I) \rangle$ . For each combination of the line parameters, 10,000 independent realizations of the same line were produced and fitted; the main results are summarized as follows:

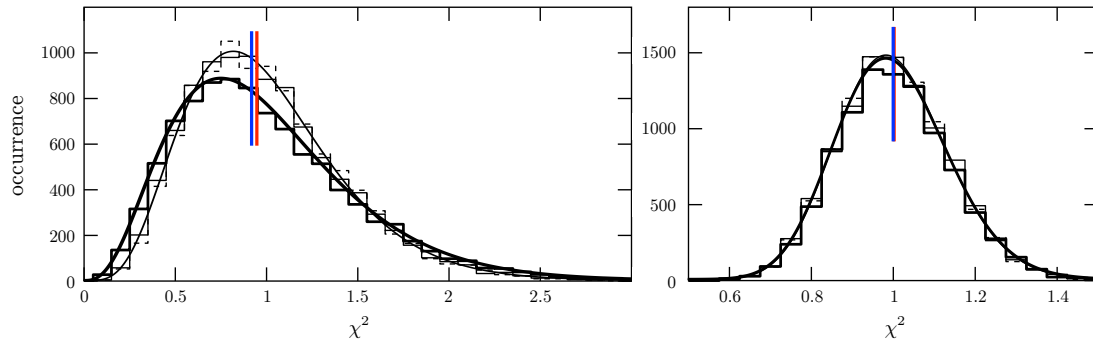


Figure 6. Value of  $\chi^2_r$  computed on 10,000 realizations of a simulated Lyman- $\alpha$  absorption line ( $\log N = 13.5$ ,  $b = 20$ ), with  $\text{SNR} = 30$ . The left and right panels show the results for a rebinned and a non-rebinned line profile, respectively. Though close to a theoretical  $\chi^2_r$  distribution (bold line), the reduced  $\chi^2_r$  computed for the rebinned profile (bold histogram) peak at a higher value; the two medians also show a statistically significant discrepancy (blue and red vertical bars, respectively). The same is not observed for the non-rebinned profile (same conventions), which has a different shape due to the higher number of data points it contains. When the  $\chi^2$  is computed with fixed line parameters, i.e. without fitting them (narrow histogram) or with  $N = 0$ ,  $b = 0$  (dashed histogram), the distribution is closer to a theoretical one (narrow line), indicating that line parameter degeneracy also plays a role in perturbing the distribution. The difference between the bold and narrow lines depend on the changing number of degrees of freedom when  $N$  and  $b$  are either fixed or fitted; this difference is totally negligible in the non-rebinned case.

- $N$  and  $b$  are correctly reconstructed in the range of  $N$  where lines are not saturated (the so-called linear regime, corresponding for a Lyman- $\alpha$  line to  $12 \lesssim \log N \lesssim 14.5$ ). At higher  $N$  values, the two parameters are degenerate and their best-fit values become strongly correlated: this as a known intrinsic effect of Voigt-profile modeling.
- On average, the line fitting procedure is not remarkably affected by the SNR, as long as the lines are distinguishable from noise. A decrease in the SNR induce a larger scatter in the distribution of the best-fit line parameters, but no bias is observed if  $N$  is above a given threshold (for a Lyman- $\alpha$  line at  $\text{SNR} \approx 30$ , the threshold is typically  $\log N \gtrsim 11$ ).
- The  $\chi^2$  test is a reliable goodness-of-fit estimator only if it is applied to non-rebinned data. When fitting is performed on the rebinned line profile, the distribution of the reduced  $\chi^2_r$  values obtained for the line realizations is generally different from an actual  $\chi^2_r$  distribution, and peaks at a higher value due to error underestimation (Fig. 6.).

The last result confirms what already discussed in Sec. 3.1. In the line fitting algorithm, the coadded spectrum is used only in the iterative phase to save computational time in finding the best-fit line parameters; those parameters are then validated on the *combined* spectrum instead, using all the non-rebinned data from the original exposures. This approach is convenient also in case of parameter degeneracy: even when the number of degrees of freedom is ill-defined, the value of reduced  $\chi^2$  computed from the combined spectrum is much better constrained by the large number of data points it contains. Tests with saturated and damped lines show that the output distribution for a rebinned line profile further deviates from an actual  $\chi^2_r$  when  $N$  increases out of the linear regime; the same is not observed for the non-rebinned profile (Fig. 6).

#### 4. PERSPECTIVES AND CONCLUSION

The ESPRESSO Data Analysis Software (DAS) encompasses a wide range of operations for both stellar and quasar spectra, which complement the standard reduction procedure to enforce a new integrated approach to the treatment of spectroscopical data. As of May 2016, seven major internal version of the software have been released and positively reviewed by ESO. All basic operation have been coded and a large portion of the interactive interface (including the workflow for the quasar branch pipeline) has been formatted. The algorithm



are currently being tested on HARPS data (for the stellar part) and on VLT UVES and VLT X-shooter data (for the quasar part) with the support of the ESPRESSO science team. Development will continue in the next months to culminate at the instrument commissioning, after which the software will be publicly released.

The significance of the DAS algorithms described in this article obviously extends beyond its application to the ESPRESSO data, and is aimed to be at the core of a new High-level Data Analysis Library, to be possibly maintained by ESO alongside its High-level Data Reduction Library.<sup>36</sup> The DAS also provides a model for the next-generation data analysis tools, such the one envisioned for the E-ELT HIRES spectrograph.<sup>37</sup> In the emerging era of precision spectroscopy, this is the only approach to ensure a validated treatment of the observational data, which is strongly called for by the increasingly ambitious science cases to be addressed.

## REFERENCES

- [1] Pepe, F., Cristiani, S., Rebolo, R., Santos, N. C., Dekker, H., Mégevand, D., Zerbi, F. M., Cabral, A., Molaro, P., Di Marcantonio, P., Abreu, M., Affolter, M., Aliverti, M., Allende Prieto, C., Amate, M., Avila, G., Baldini, V., Bristow, P., Broeg, C., Cirami, R., Coelho, J., Conconi, P., Coretti, I., Cupani, G., D’Odorico, V., De Caprio, V., Delabre, B., Dorn, R., Figueira, P., Fragoso, A., Galeotta, S., Genolet, L., Gomes, R., González Hernández, J. I., Hughes, I., Iwert, O., Kerber, F., Landoni, M., Lizon, J.-L., Lovis, C., Maire, C., Mannetta, M., Martins, C., Monteiro, M. A., Oliveira, A., Poretti, E., Rasilla, J. L., Riva, M., Santana Tschudi, S., Santos, P., Sosnowska, D., Sousa, S., Spanò, P., Tenegi, F., Toso, G., Vanzella, E., Viel, M., and Zapatero Osorio, M. R., “ESPRESSO – An Echelle SPectrograph for Rocky Exoplanets Search and Stable Spectroscopic Observations,” *The Messenger* **153**, 6–16 (2013).
- [2] Mayor, M., Pepe, F., Queloz, D., Bouchy, F., Rupprecht, G., Lo Curto, G., Avila, G., Benz, W., Bertaux, J.-L., Bonfils, X., Dall, T., Dekker, H., Delabre, B., Eckert, W., Fleury, M., Gilliotte, A., Gojak, D., Guzman, J. C., Kohler, D., Lizon, J.-L., Longinotti, A., Lovis, C., Megevand, D., Pasquini, L., Reyes, J., Sivan, J.-P., Sosnowska, D., Soto, R., Udry, S., van Kesteren, A., Weber, L., and Weilenmann, U., “Setting New Standards with HARPS,” *The Messenger* **114**, 20–24 (2003).
- [3] Cosentino, R., Lovis, C., Pepe, F., Collier Cameron, A., Latham, D. W., Molinari, E., Udry, S., Bezawada, N., Black, M., Born, A., Buchschacher, N., Charbonneau, D., Figueira, P., Fleury, M., Galli, A., Gallie, A., Gao, X., Ghedina, A., Gonzalez, C., Gonzalez, M., Guerra, J., Henry, D., Horne, K., Hughes, I., Kelly, D., Lodi, M., Lunney, D., Maire, C., Mayor, M., Micela, G., Ordway, M. P., Peacock, J., Phillips, D., Piotto, G., Pollacco, D., Queloz, D., Rice, K., Riverol, C., Riverol, L., San Juan, J., Sasselov, D., Segransan, D., Sozzetti, A., Sosnowska, D., Stobie, B., Szentgyorgyi, A., Vick, A., and Weber, L., “HARPS-N: the new planet hunter at TNG,” *Proc. SPIE* **8446**, 84461V (2012).
- [4] Webb, J. K., King, J. A., Murphy, M. T., Flambaum, V. V., Carswell, R. F., and Bainbridge, M. B., “Indications of a Spatial Variation of the Fine Structure Constant,” *Physical Review Letters* **107**(19), 191101 (2011).
- [5] King, J. A., *Searching for variations in the fine-structure constant and the proton-to-electron mass ratio using quasar absorption lines*, PhD thesis, PhD Thesis, 2012 (2012).
- [6] Rahmani, H., Wendt, M., Srianand, R., Noterdaeme, P., Petitjean, P., Molaro, P., Whitmore, J. B., Murphy, M. T., Centurion, M., Fathivavari, H., D’Odorico, S., Evans, T. M., Levshakov, S. A., Lopez, S., Martins, C. J. A. P., Reimers, D., and Vladilo, G., “The UVES large program for testing fundamental physics - II. Constraints on a change in  $\mu$  towards quasar HE 0027-1836,” *MNRAS* **435**, 861–878 (2013).
- [7] Marconi, A. et al., “HIRES the high-resolution spectrograph for the E-ELT,” *Proc. SPIE* (2016).
- [8] Sandage, A., “The Change of Redshift and Apparent Luminosity of Galaxies due to the Deceleration of Selected Expanding Universes,” *ApJ* **136**, 319 (1962).
- [9] Liske, J., Grazian, A., Vanzella, E., Dessauges, M., Viel, M., Pasquini, L., Haehnelt, M., Cristiani, S., Pepe, F., Avila, G., Bonifacio, P., Bouchy, F., Dekker, H., Delabre, B., D’Odorico, S., D’Odorico, V., Levshakov, S., Lovis, C., Mayor, M., Molaro, P., Moscardini, L., Murphy, M. T., Queloz, D., Shaver, P., Udry, S., Wiklund, T., and Zucker, S., “Cosmic dynamics in the era of Extremely Large Telescopes,” *MNRAS* **386**, 1192–1218 (2008).

- [10] McKay, D. J., Ballester, P., Banse, K., Izzo, C., Jung, Y., Kiesgen, M., Kornweibel, N., Lundin, L. K., Modigliani, A., Palsa, R. M., and Sabet, C., “The common pipeline library: standardizing pipeline processing,” *Proc. SPIE* **5493**, 444–452 (2004).
- [11] Ballester, P., Modigliani, A., Boitquin, O., Cristiani, S., Hanuschik, R., Kaufer, A., and Wolf, S., “The UVES Data Reduction Pipeline,” *The Messenger* **101**, 31–36 (2000).
- [12] Modigliani, A., Mulas, G., Porceddu, I., Wolff, B., Damiani, F., and Banse, B. K., “The FLAMES-UVES Pipeline,” *The Messenger* **118**, 8–10 (2004).
- [13] Modigliani, A., Goldoni, P., Royer, F., Haigron, R., Guglielmi, L., François, P., Horrobin, M., Bristow, P., Vernet, J., Moehler, S., Kerber, F., Ballester, P., Mason, E., and Christensen, L., “The X-shooter pipeline,” *Proc. SPIE* **7737**, 773728 (2010).
- [14] Warmels, R. H., “The ESO-MIDAS System,” *Astronomical Society of the Pacific Conference Series* **25**, 115 (1992).
- [15] Di Marcantonio, P., Sosnowska, D., Lovis, C., D’Odorico, V., Cupani, G., Sousa, S., González Hernández, J. I., Cirami, R., Lo Curto, G., Mégevand, D., and Cristiani, S., “Challenges and peculiarities of ESPRESSO data flow cycle: from target choice to scientific results,” *Proc. SPIE* **8448**, 84481O (2012).
- [16] Di Marcantonio, P., D’Odorico, V., Cupani, G., Sosnowska, D., Lovis, C., Sousa, S., Figueira, P., González Hernández, J. I., Lo Curto, G., Modigliani, A., Cirami, R., Mégevand, D., and Cristiani, S., “ESPRESSO data flow: from design to development,” *Proc. SPIE* **9149**, 91491Q (2014).
- [17] Cupani, G., Cristiani, S., D’Odorico, V., González-Hernández, J. I., Lovis, C., Segovia Milla, A. G., Sousa, S., Sosnowska, D., Di Marcantonio, P., and Mégevand, D., “Data treatment towards the ELT age. The ESPRESSO case,” *Mem. Società Astronomica Italiana* **86**, 502 (2015).
- [18] Cupani, G., D’Odorico, V., Cristiani, S., González-Hernández, J., Lovis, C., Sousa, S., Vanzella, E., Marcantonio, P. D., and Mégevand, D., “Data Analysis Software for the ESPRESSO Science Machine,” *Astronomical Society of the Pacific Conference Series* **495**, 289 (2015).
- [19] Freudling, W., Romaniello, M., Bramich, D. M., Ballester, P., Forchi, V., García-Dabó, C. E., Moehler, S., and Neeser, M. J., “Automated data reduction workflows for astronomy. The ESO Reflex environment,” *A&A* **559**, A96 (2013).
- [20] Rudie, G. C., Steidel, C. C., Shapley, A. E., and Pettini, M., “The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift  $\langle z \rangle = 2.4$ ,” *ApJ* **769**, 146 (2013).
- [21] Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., and Monteiro, M. J. P. F. G., “A new code for automatic determination of equivalent widths: Automatic Routine for line Equivalent widths in stellar Spectra (ARES),” *A&A* **469**, 783–791 (2007).
- [22] Sousa, S. G., Santos, N. C., Adibekyan, V., Delgado-Mena, E., and Israelian, G., “ARES v2: new features and improved performance,” *A&A* **577**, A67 (2015).
- [23] Bonifacio, P., “From observations to abundances: equivalent widths and line profile fitting,” *Memorie della Società Astronomica Italiana Supplementi* **8**, 114 (2005).
- [24] Kelson, D. D., “Optimal Techniques in Two-dimensional Spectroscopy: Background Subtraction for the 21st Century,” *The Publications of the Astronomical Society of the Pacific* **115**, 688–699 (2003).
- [25] Fruchter, A. S. and Hook, R. N., “Drizzle: A Method for the Linear Reconstruction of Undersampled Images,” *The Publications of the Astronomical Society of the Pacific* **114**, 144–152 (2002).
- [26] Galassi, M. et al., [*GNU Scientific Library Reference Manual - Third Edition*], RRP (2009).
- [27] Gray, D. F., “Spectral line-depth ratios as temperature indicators for cool stars,” *PASP* **106**, 1248–1257 (1994).
- [28] Sousa, S. G., Alapini, A., Israelian, G., and Santos, N. C., “An effective temperature calibration for solar type stars using equivalent width ratios. A fast and easy spectroscopic temperature estimation,” *A&A* **512**, A13 (2010).
- [29] Sousa, S. G., Santos, N. C., and Israelian, G., “TMCalc - a fast code to derive  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  for FGK stars,” *A&A* **544**, A122 (2012).
- [30] Gray, D. F., [*The Observation and Analysis of Stellar Photospheres*] (2005).

- [31] Castelli, F. and Kurucz, R. L., “New Grids of ATLAS9 Model Atmospheres,” *IAU Symposium* **210**, A20 (2003).
- [32] Pâris, I., Petitjean, P., Rollinde, E., Aubourg, E., Busca, N., Charlassier, R., Delubac, T., Hamilton, J.-C., Le Goff, J.-M., Palanque-Delabrouille, N., Peirani, S., Pichon, C., Rich, J., Vargas-Magaña, M., and Yèche, C., “A principal component analysis of quasar UV spectra at  $z \sim 3$ ,” *A&A* **530**, A50 (2011).
- [33] Sánchez-Ramírez, R., Ellison, S. L., Prochaska, J. X., Berg, T. A. M., López, S., D’Odorico, V., Becker, G. D., Christensen, L., Cupani, G., Denney, K. D., Pâris, I., Worseck, G., and Gorosabel, J., “The evolution of neutral gas in damped Lyman  $\alpha$  systems from the XQ-100 survey,” *MNRAS* **456**, 4488–4505 (2016).
- [34] Stevans, M. L., Shull, J. M., Danforth, C. W., and Tilton, E. M., “HST-COS Observations of AGNs. II. Extended Survey of Ultraviolet Composite Spectra from 159 Active Galactic Nuclei,” *ApJ* **794**, 75 (2014).
- [35] Aaronson, M., McKee, C. F., and Weisheit, J. C., “The identification of absorption redshift systems in quasar spectra,” *ApJ* **198**, 13–30 (1975).
- [36] Ballester, P., Gabasch, A., Jung, Y., Modigliani, A., Taylor, J., Coccato, L., Freudling, W., Neeser, M., and Marchetti, E., “The High Level Data Reduction Library,” *Astronomical Society of the Pacific Conference Series* **495**, 383 (2015).
- [37] Cupani, G., D’Odorico, V., Cristiani, S., González Hernández, J. I., Lovis, C., Sousa, S., Calderone, G., R., C., Di Marcantonio, P., and Mégevand, D., “E-ELT HIRES the high resolution spectrograph for the E-ELT: integrated data flow system,” *Proc. SPIE* (2016).