Mem. S.A.It. Vol. 86, 502
© SAIt 2015

Memorie della

Data treatment towards the ELT age

The ESPRESSO case

G. Cupani[1], S. Cristiani[1], V. D'Odorico[1], J. I. González-Hernández[2],
C. Lovis[3], A. G. Segovia Milla[3], S. Sousa[4], D. Sosnowska[3],
P. Di Marcantonio[1], and D. Mégevand[3]

[1] Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy, e-mail: `cupani@oats.inaf.it`
[2] Instituto de Astrofísica de Canarias, Vía Láctea, E-38205 La Laguna, Tenerife, Spain
[3] Université de Genève, 51 Chemin des Maillettes, CH-1290 Versoix, Switzerland
[4] Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, PT-4150-762 Porto, Portugal

Abstract. Several ambitious scientific projects are currently pushing the limits of astrophysical spectroscopy, in terms of wavelength accuracy and long-term stability. These objectives put strong constraints also on the treatment of observational data, requiring dedicated software to be developed as an integral part of the instrument. In this context, a key role will be played by ESPRESSO, an ultra-stable, high-resolution spectrograph for the VLT. ESPRESSO will be the first ESO instrument to be delivered with a dedicated tool for data analysis, in addition to data reduction. New solutions to treat ESPRESSO data have been developed in view of their application to the the next-generation spectrographs.

Key words. Instrumentation: spectrographs – Methods: data analysis – Cosmology: observations

1. Introduction

In the recent years, observational cosmology is gradually evolving into a precision science. A number of compelling projects depend on the possibility to observe with unprecedented accuracy the absorption features imprinted on the spectra of distant sources (e.g. quasars) by the intervening inter-galactic medium (IGM). A description of the current scenario can be found in another contribution to this volume (Cristiani et al. 2015); we recall here only the most relevant science cases in the field of fundamental physics. These are: (i) the measurement of a possible variability of the dimensionless constants α (fine-structure constant) and μ (proton-to-electron mass ratio) across the cosmic time (detected from a redshift-dependent shift between associated absorption lines: see e.g. Webb et al. 2011, King 2012, Rahmani et al. 2013), and (ii) the direct observation of the expansion of the Universe from the determination of a time-drift in the redshift of absorbers, the so-called Sandage test (Liske et al. 2008).

Both these science cases pose extreme technological challenges, as they require high sensitivity (to achieve a significant signalto-noise ratio with $z > 2$ sources), high resolution over a wide wavelength range (to resolve and correlate distant absorption lines), and an extreme thermo-mechanical stability. The required wavelength accuracy is ~ 10 cm s^{-1} for the fundamental constant variability and ~ 1 cm s^{-1} (to be maintained across at least two decades) for the Sandage test. To achieve these limits, specifically-designed instruments have been conceived. One such instrument is ESPRESSO (Pepe et al. 2013), the Echelle SPectrograph for Rocky Exoplanets And Stable Spectral Observations, currently under construction for the ESO Very Large Telescope.

In this article we discuss the case of ESPRESSO as a forerunner to the high-resolution spectrograph envisioned for the European Extremely Large Telescope (E-ELT) (Maiolino et al. 2013; Oliva et al. 2015). We will focus in particular on the scientific software which is meant to complement both instruments to fully exploit their capabilities. A new science-driven approach to instrument design is described, the so-called "science machine" concept, which regards the data treatment as an integral part of the spectrograph. The key idea is that the exceptional requirements raised by the current age of high-precision cosmology cannot be met without implementing *ad-hoc* solutions to extract scientific information from the data.

2. ESPRESSO

ESPRESSO is an ultra-stable, fibre-fed echelle spectrograph which combines and enhances the capabilities of UVES (Dekker et al. 2000), HARPS (Mayor et al. 2003), and HARPS-N (Cosentino et al. 2012). Its first light is foreseen for 2017. It will operate at very high (130,000 to 200,000) and high (55,000) resolution using respectively one or all four VLT Unit Telescopes (UT), thanks to its location in the coudé combined laboratory of Paranal. In the 4-UT mode, the equivalent collecting area will be the same of a 16 meter telescope; in the 1-UT mode, the wavelength accuracy will be better than 10 cm s^{-1}. All these specifications are meant to bridge the gap towards E-ELT high-resolution spectrograph, which will profit from a 39-meter telescope collecting area and will reach an accuracy of 1 cm s^{-1}, using similar solutions (vacuum vessels and thermal chambers) to maintain stability.

ESPRESSO's main science cases are the search of Earth-like exoplanets in the habitable zone and the study of fundamental constant variability; as a consequence, it will observe mainly stars and quasars. This very consistent set of targets strongly favours the development of a dedicated software suite covering all the requirements of data treatment. In fact, ESPRESSO is the first VLT instrument to include among its deliverables a tool to analyse the data, the Data Analysis Software (DAS), developed alongside the customary Data Reduction Software (DRS). The first complete release of both packages is foreseen for the end of 2015.

Both the DRS and the DAS are designed and written in compliance with the ESO Data Flow System, which manages all the steps from observation preparation to data treatment and archiving. They consist of independent ANSI-C plug-ins ("recipes") based on the ESO Common Pipeline Library (CPL; McKay et al. 2004), each one performing a separate operation. The plug-ins can be executed as stand-alone modules or in cascade ("pipeline"); the DRS recipes, in particular, will be triggered on the fly by the injection of raw data as soon as they are produced by the instrument, while the DAS will be called at the user's discretion. In this respect, ESPRESSO will work as a real "machine to produce science", as the observers will be able to obtain high-level science products from the instrument (e.g. radial velocities of stars; accurate wavelengths, column densities, and Doppler widths of the absorbers towards distant quasars) within minutes after the execution of the observing blocks. The DAS is particularly relevant in this framework as it allows to close the data flow cycle, by driving the preparation of new observations depending on the scientific output of the previous ones. This model will naturally fit in the E-ELT high-resolution spectrograph concept.

2.1. Data Reduction Software

The DRS includes 12 recipes, responsible for (i) creating a master bias, a master dark, and a master flat; (ii) determining the gain, the pixel linearity and the location of bad and hot pixels; (iii) determining the position and geometry of the orders, the geometry of pixels, and the wavelength calibration; (iv) computing the relative efficiency of the fibres, removing contamination and performing flux calibration; and (v) extracting the spectrum of the science target.

The most innovative calibration source of ESPRESSO is the Laser-Frequency Comb (LFC; Wilken et al. 2012, Lo Curto et al. 2012), which consists of a regular array of laser-generated emission lines spanning the whole wavelength range of the instrument, with a foreseen accuracy is $\Delta\lambda/\lambda \sim 10^{-10}$. The LFC output will be available as a simultaneous reference spectrum on the same frame of the science observations; besides for wavelength calibration, it can be used to monitor any irregularities in size and sensitivity of the detector pixel grid. The assumption that all pixels have equal size can lead to systematic errors of tens of m s^{-1} (Molaro et al. 2013). Tests on HARPS-N show that the LFC is capable of determining relative variations in pixel size of $\sim 10^{-4}$, whose effect would be unrecoverable by flat-fielding alone. The DRS will monitor and correct for such irregularities.

Another relevant feature of the ESPRESSO approach to data reduction is the conservative treatment of detector readings (i.e. photon counts and relative errors). It is quite common, when reducing echelle spectra with curved order, to "rectify" them into a single merged spectrum whose pixels do not correspond one-to-one to the original detector pixels. This process, called "rebinning", is performed during wavelength calibration by combining the contribution of different pixels, and introduces a spurious correlation in the flux errors (Bonifacio 2005). Under such circumstances, standard best-fitting techniques like the χ^2 test, although routinely applied, cannot be safely used to interpret the spectral features. On the contrary, the DRS will propagate the detector pixels throughout the reduction chain with their original wavelengths, as determined by the LFC. Along with the standard rebinned spectrum, a set of non-rebinned 2-d matrices will be issued, containing the flux values, the relative errors, and the wavelengths of all pixels in the different echelle orders.

2.2. Data Analysis Software

The DAS includes of 13 recipes, split into two branches, for the analysis of stellar and quasar spectra respectively (Cupani et al. 2015). The stellar recipes are responsible for (i) measuring the radial velocity of the star, either with the cross-correlation method or by comparison with synthetic spectra; (ii) computing the stellar activity indexes; (iii) estimating the stellar continuum and comparing observed spectra with synthetic spectra; (iv) measuring the equivalent width of absorption lines and estimating of the effective temperature and [Fe/H] metallicity (only for FGK stars). The quasar recipes are responsible for (i) determining the level of the quasar continuum emission; (ii) fitting Voigt profiles to the absorption lines; and (iii) identifying the absorption systems, assigning to each line the ionic transition that most probably produced it. A set of common utilities have also been developed to combine different exposures of the same object, to mask spectral regions, and to detect the absorption lines.

The DAS was designed to combine the somehow conflicting needs of automation, control, and flexibility. Unlike data reduction, data analysis is not a one-way procedure. Users typically refine the results through multiple iterations; in the case of quasar spectra analysis, in particular, each task takes advantage from the results of the others. This model will be implemented in the software by exploiting the capabilities of the Reflex enviroment (Freudling et al. 2013), which is currently offered by ESO as the primary interface to the instrument pipelines. Reflex is developed within the Kepler workflow engine and takes care of organizing the I/O data, executing the recipes, and interactively displaying the results. Using Python scripts, we have devised a way to feed into the workflow the information extracted from its previous executions.

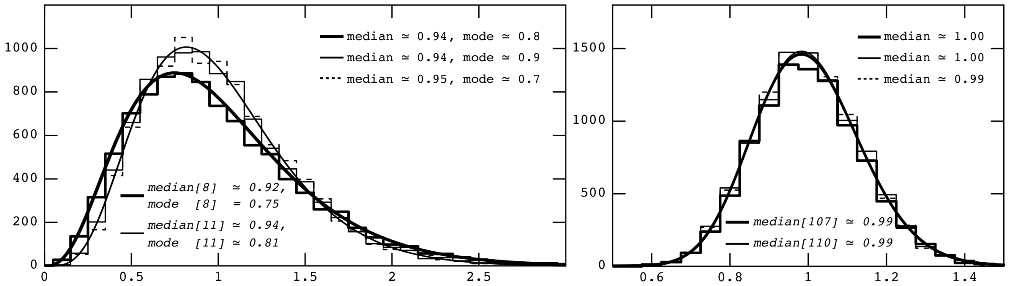

Fig. 1. Value of χ^2 computed on 10,000 realisation of a simulated absorption line. Left panel: results with a rebinned spectrum obtained combining 10 simulated exposures (11 pixels); right panel: results with the collective spectrum of all the exposures (110 pixels). Three tests were performed: (a) we run the code to determine the best-fitting line parameters (i.e. position, column density, Doppler width), (b) we compared the individual realisations with the known theoretical profile, and (c) we run the code on a null line (i.e. on a spectrum of un-absorbed, normalized continuum). The degrees of freedom in the latter two cases correspond to the number of pixels, while in the first case this value is decreased by the number of fitted parameters. The obtained distributions of reduced χ^2 values (bold, thin, and thin dashed histograms, for (a), (b), and (c) respectively) are superimposed to the expected theoretical distributions (bold and thin lines); the relevant location parameters of the distributions are also given (theoretical values in italic).

Algorithmically, the DAS implements several new solutions. An iterative technique for estimating the level of quasar continuum emission (Cupani et al. 2015) has been implemented, relying on a progressive removal of all the absorption features (modeled with Voigt profiles). The Voigt-fitting module employs the package `gsl_multifit` from the GNU Scientific Library (GSL, Galassi et al. 2009) and applies the χ^2-test on the non-rebinned 2-d spectra produced by the DRS.

The reliability of the fitting procedure has been widely tested on simulated absorption line; some results are shown in Fig. 1. As expected, the fit is not well constrained when performed on rebinned spectra (left panel): the median of distribution of χ^2 values obtained over 10,000 different realizations departs from the expected values, partly because the errors are underestimated (due to rebinning) and partly because the degrees of freedom (d.o.f.) are ill-defined (since the line parameters are correlated). The problem can be overcome by constraining the fit on the non-rebinned spectra (right panel): in this case no spurious correlation is introduced, and the d.o.f. are enough to cancel any uncertainty on the expected shape of the distribution. Such careful treatment of the data is of paramount importance in the incoming age of high-precision spectroscopy.

References

Bonifacio, P. 2005, MmSAI, 8, 114
Cosentino, R., et al. 2012, Proc. SPIE, 8446, 84461V
Cristiani, S., et al. 2015, MmSAI, 86, 486
Cupani, G., et al. 2015, PASP (in press)
Dekker, H., et al. 2000, Proc. SPIE, 4008, 534
Di Marcantonio, P., et al. 2014, Proc. SPIE, 9149, 91491Q
Freudling, W., et al. 2013, A&A, 559, A96
Galassi, M. et al. 2009, GNU Scientific Library Reference Manual
King, J. A. 2012, `arxiv:1202.6365`
Liske, J., et al. 2008, MNRAS, 386, 1192
Lo Curto, et al. 2012, The Messenger, 149, 2
Maiolino, R., et al. 2013, `arxiv:1310.3163`
Mayor, M., et al. 2003, The Messenger, 114, 20
McKay. D. J., et al. 2004, Proc. SPIE, 5493, 444
Molaro, P., et al. 2013, A&A, 560, 61
Oliva, E., et al. 2015, MmSAI, 86, 474
Pepe, F., et al. 2013, ESO Messenger, 153, 6
Rahmani H., et al. 2013, MNRAS, 435, 861
Webb, J., et al. 2011, Phys. Rev. Lett., 107, 191101
Wilken, T., et al. 2012, Nature, 485, 611