The Data-Reduction Software for GIRAFFE, the VLT medium resolution multi-object fiber-fed spectrograph

André Blecha[a], Véronique Cayatte[b], Pierre North[c], Frédéric Royer[a] and Gilles Simond[a]

[a]Observatoire de Genève, 51, ch. des Maillettes, CH-1290 Sauverny, Switzerland
[b]DAEC, Observatoire de Paris-Meudon, 5, Place Jules Janssen, F-92195 Meudon cedex, France
[c]Institut d'astronomie de l'Universite de Lausanne, CH-1290 Chavannes-des-Bois, Switzerland

ABSTRACT

A brief overview is given of the Data-Reduction Software (DRS) being developed at Geneva and Paris Observatories for the fiber-fed multi-object GIRAFFE spectrograph - part of the VLT FLAMES facility. The specific aspects of each of three modes (MEDUSA: single-fiber/object, IFU: 20 fibers/object, ARGUS: all fibers on one object) is presented and discussed. The localization and the extraction are described in more details, the original features are outlined and the critical issue addressed. We discuss the problem of the PSF variation across the surface of the detector and the impact on the sky subtraction. The strategy of the real-time data quality assessment using the simultaneous calibration exposure is described. Some aspect of the DRS implementation and the VLT environment are discussed. The *dirty prototype* of the DRS under Matlab including the crude instrument simulator is discussed and some experiments measuring the expected accuracy of radial-velocity are shown.

Keywords: Multi-object spectroscopy, data-reduction, fiber-fed spectroscopy, simultaneous calibration, radial velocity

1. INTRODUCTION

1.1. Framework of the project

The GIRAFFE spectrograph is part of the FLAMES[1] fiber-fed facility for Unit Telescope #2 (UT2) of VLT and is due for operation by the fall 2001. The GIRAFFE instrument is built by the consortium formed by ESO, Observatoire de *Paris-Meudon* in France (OP) and the Swiss entity referred hereafter as OGL including the *Observatoire de Genève* and the *Institut d'astronomie de l'Université de Lausanne*. The development of the Data Reduction Software (DRS) is made under the contract between the ESO and the OGL/OP groups while the overall instrument development is directly under the control of ESO. The high-level software developed for FLAMES could be divided in three main packages:

1. The Observing and Preparation Software (OPS)

   It includes all pieces necessary to prepare and carry on the standard setup of all sub-assemblies of the facility as well as the data acquisition (observation, calibration and maintenance). It generates the raw data handled by the DRS.

2. The Data Reduction Software (DRS)

   It encompasses the software necessary to process the raw data in order to produce all necessary *clean* calibration data (to build the instrument calibration database) and a set of flux-and-wavelength-calibrated spectra in all standard observing modes. The DRS will exist in an *on-line* version where modules are executed unattended within the VLT Data Flow System (DFS) and in an *off-line* version with modules executed within the standard environment in an automatic or interactive mode.

3. The Data Analysis Software (DAS)

   The data-analysis is not included in the DRS. However, some analysis software implementing few instrument specific functions not routinely available or not at the appropriate level within the commonly used data analysis systems will be developed. Only an off-line version will be available (no constraints of integration into the DFS) and will not be dependent on a specific data-analysis system.

Send correspondence to Andre.Blecha@obs.unige.ch

1.2. Instrument overview

In the following we describe the GIRAFFE observing modes and we discuss the specific aspects of the data produced under each observing mode. Three observing modes are available:

1. In the MEDUSA mode 132 fibers are distributed in a field of 25 arcmin diameter. Some of these have to be attributed to sky measurement.

   Since the individual objects are not, in general, physically related, there is no correlation between the individual spectra and the contrast between two adjacent spectra on the detector could be very high. On the other hand the extent of the fiber image on the detector is large enough to give a comfortable description of the local PSF and the large separation between spectra prevents a significant overlap.

2. In the IFU (Integral Field Units) mode, 15 fiber bundles are distributed in the same 25 arcmin field, each bundle containing 20 fibers and corresponding to an area of 3×2 arsec on the sky (spatial sampling of 0.52 arcsec). In addition, the sky is measured at 15 positions in the field of view by dedicated fibers.

   In this mode a substantial degree of correlation exists between adjacent image elements at the focal plane, both because of the PSF sampling and because the extended objects present some spatial continuity of their energy distribution. Therefore there should be no dramatic spectrum-to-spectrum contrast in the IFU and ARGUS modes. On the other hand the PSF on the detector is much more difficult to disentangle because it is less well sampled (the fiber size being smaller) and the spectra undergo significant cross-talk (being close-packed).

3. ARGUS, a single integral field unit containing 300 fibers covering a rectangular area of 11.5×7.3 arcsec (spatial sampling of 0.52 arcsec) or 6.6×4.2 arcsec (spatial sampling of 0.3 arcsec). As for the IFU mode, 15 sky fibers are available distributed anywhere around the Argus field (the microlens aperture on the sky is 0.52×0.52 arcsec).

   As to the data, the case could be assimilated to a single IFU excepted perhaps for the sky modeling.

For each of these modes, two resolving powers are offered: a low resolution mode with $R \sim 5000$ and $R \sim 8000$ in the MEDUSA and IFU/Argus modes respectively, and a high resolution mode with $R \sim 15000$, respectively $R \sim 24000$. The better spectral resolution in the IFU/Argus mode is due to the better sampling on the sky (the smaller diameter of fibers).

2. SPECIFIC FEATURES OF THE GIRAFFE INSTRUMENT

Unlike the UVES coupling to the fiber positioner that carries light to an existing spectrograph not initially designed to that purpose, the GIRAFFE spectrograph is fully designed as a fiber-coupled instrument and integrates some original features. The most outstanding is the simultaneous calibration using the Th-Ar and Ne lamps. Though the primary motivation to commit permanently 5 spectra on each exposure to that purpose was the requirement of high accuracy of the wavelength solution, the by-product of this situation - the presence, in any mode, of five optimally exposed emission spectra that are close-to-uniformly distributed over the detector - is a substantial asset. At least 20 usable lines/spectrum are available in each spectral band. The number of lines and the scrambling effect of the fibers insure that the PSF variation both in form and position could be efficiently modeled for every single exposure. This is an exceptional opportunity to monitor most of instrumental parameters and greatly influences the design of the data reduction software since it may heavily rely on a very accurate localization.

It is generally considered that sky subtraction is the weak point of fiber spectrograph (relative to MOS spectrographs with slitlets punched on masks), essentially because the sky cannot be measured at the very position of each object and because uncertainties in both fiber transmission and PSF variation hamper a precise sky subtraction. In the case of GIRAFFE, the first limitation evidently remains, but an important effort will be made to overcome the second one. In particular, the relative fiber transmission will be measured to better than 2% through standard calibrations, the fibers being arranged in a special configuration on the field-plate. Higher precision cannot be guaranteed with this method because the fibers will be re-configured between calibration and observations, and the field-plate moved from the positionner to the Nasmyth focal plane, which will change the fiber tension and, as a result, slightly change fiber transmission. An illuminated Nasmyth screen will be provided in order to overcome this problem, enabling the observer to measure the relative fiber transmission in the science configuration, just after or before the science exposure.

The problem of PSF variation on the surface of the detector is addressed in Section 4.3.

3. THE DRS OVERVIEW

Traditionally, the DRS is an off-line package and though it will be used within the framework of the VLT acquisition process, the on-line aspect has only little impact on the DRS requirements. The configurable level of interactivity going from unattended automatic mode during the standard acquisition to full interactivity during the instrument maintenance runs and robustness required for a fully automatic operation are the main features that distinguish the GIRAFFE DRS from the classical off-line approach.

3.1. Interfaces

3.1.1. Inputs

On the input, the DRS receives the raw images from the VLT acquisition pipeline, the master calibration data and the calibration constants extracted from the instrument database and the control parameters provided through configuration files.

There is no difference, except for the raw images and the control parameters access, between the use of the DRS within the VLT DFS pipeline or off-line.

3.1.2. Outputs

The DRS produces the following data:

- Flux and wavelength calibrated spectra with optionally error and number of pixels in each bin,
- Standards set of numerical values that characterize the results and are used for the long term performance monitoring and quality control,
- Set of report/history files and graphics;

Optionally all necessary calibration and control data can be exported for further off-line use.

3.2. Main components

- Data access layer (DAL)

  This layer is included in order to make the DRS independent of the environment and of the access to the instrument database. The DRS works only with standard and simple data structures (FITS for images and ASCII for tables and configuration files).

- The Processing module (PM)

  This software provides all computation tools necessary to carry out the full data reduction. It could be divided in:

  - Preprocessing
  - Localization
  - Extraction
  - Wavelength calibration and rebinning
  - Global Flat-Fielding
  - Sky modeling and subtraction
  - Flux spectrophotometric calibration
  - Image reconstruction

  Note that PM includes the tools necessary to the quality control and performance assessment.

- The graphics and image displays (GID)

  Beside the standard components such as classical graphical tools and image displays which are available in the VLT Common Software as well as in many public-domain environments, there are components specific to the intrinsic 3-D nature of the data to be displayed such as the Graphical field navigator. This tool displays the field according to the information provided by the input data (object description in MEDUSA and IFU mode, the integrated image in ARGUS mode with optionally in both cases the field image in the background) and enables display functions on part or totality of data selected graphically.

- The graphical user interface (GUI) both for scientific and maintenance mode.

- The External Components.

  This includes all public domain or VLT standard software used within the DRS.

3.3. Quality control

Virtually all high level functions return parameters that are used for the quality control. They are described in the detailed functional specifications ([3]).

The quality control level 0 (QC0) is a part of the normal data reduction processing. This is done for all processed frames including the science frames in the DFS pipeline. With the exception of the Flat Field (FF) calibration exposures, any GIRAFFE frame includes 5 optimally exposed spectra used for the simultaneous wavelength calibration. This is the opportunity for a full check of the actual instrument performances compared to the nominal ones:

1. read-out noise and CCD offset,
2. number of cosmic hits and detector defects,
3. proportion of the scattered light,
4. spectra localization stability,
5. instrument throughput (integrated light per spectrum/calibration ,illumination)
6. wavelength solution stability and spectral resolution;

The quality control level 1 (QC1) detecting trends and producing the average certified calibration data is not a part of the DRS.

3.4. Design features

The ESO VLT framework imposes formal conditions ([2]) that prevent or make inefficient the adaptation of the existing packages at the code-level. This is the opportunity to introduce a few interesting features not common in existing astronomical software.

- Associated to the processed image, the numerical mask badpixel(x, y) of the *bad pixels* is initialized and progressively updated through the reduction process. The analysis software supports *flagged* pixels all through the analysis. The flagging is implemented in a way to identify the reason of point rejection by value of the mask, while the mask itself is used as a logical operand (points with value greater than 0 are rejected). Alternatively, the possibility to replace flagged-pixel at any stage of the data reduction is implemented.

- For each pixel of the extracted spectra, an internal error is computed, taking into account the mask of bad pixels. The software handling the extracted spectra takes optionally into the account such error.

4. DRS FUNCTIONS

Table 1 in the Appendix lists the main functions of the GIRAFFE DRS. We refer the reader to the detailed description of the functional specifications ([3]). Two very obvious facts are worth noticing. First of all, for the majority of functions, the most widespread public-domain data/image processing packages offer similar functions in specific environments. Also, when comparing requirements for the first and second generation of the VLT instruments it becomes obvious that many functions may be potentially re-used in other contexts. This is the reason why we discuss the implementation issue in section 6 even though the detailed design is yet to be made. In the following we focus on a few specific aspects of the GIRAFFE DRS.

4.1. Localization

The localization strategy is based on the excellent long-term set-up repeatability (within 3 pixels), the negligible relative random displacements of spectra (mechanical stability of the slit assembly) and the analytical model with adjustable parameters linking the position on the entrance slit to the position on the CCD detector. Such a model gives both the position and the width of the projected image of the entrance fiber on the CCD.

The localization starts from a simple optical model and operates on a complete set or a subset of the data. The same function is used to derive a complete localization solution or to adjust the current solution using 5 simultaneous wavelength calibration spectra only. The localization proceeds according the following steps:

- Using the current localization, compute the centroid and optionally the width for each spectral bin of each spectrum.
- Select bins that are used for the fit on the basis of the signal level, number of valid pixels in the bin and estimated error of the bin localization.
- Fit parameters of the localization correction model on selected bins.
- Update the current localization.

4.2. Extraction

The extraction process rests on the same technical assumption as the localization. It makes use of the parameters of the localization **with no further correction** allowed in that respect. This is possible only because the localization is adjusted on each exposure on well-exposed calibration spectra. The extraction proceeds as follows:

- By default, the pixels are weighted proportionally to the inverse of their estimated variances. Other weighting schemes are possible.
- For each spectral bin an analytical model is fitted on the multiple spectra profile (including all spectra in the spectral bin). Optionally, if the cross-talk between spectra can or is wanted to be neglected, this step could be done separately for each spectrum by Horne's method ([4]).
- The intensities, the standard deviations σ and the local backgrounds of the extracted spectra are set to the amplitudes, errors and background terms of the fitted model.

During the extraction, the flagged points are not used for the fit. The standard error σ of the flux at each pixel of the extracted spectrum is estimated and saved. This provides an independent estimate of the overall error for each spectral element to be compared to the expected shot-noise limit.

4.3. Sky modeling and subtraction

This step is crucial for faint objects. Note that the accuracy of the subtraction is fully dependent on the accuracy of the measurement and stability of the fiber transmission.

In order to be able to directly subtract an extracted sky spectrum from an extracted object spectrum, both should have the same PSF. This is not granted, essentially because the PSF, and therefore the resolution, varies according to the position on the detector in both directions.

Viton& Milliard ([5]) devised a method to get around this difficulty assuming that, for a given wavelength, the PSF could be described by a smoothly varying function of the position on the detector. In this case a polynomial model could be fitted at each spectral bin to all sky spectra and realistic sky spectra could be computed for any science spectra using the model.

The above method is not directly applicable to GIRAFFE because of the large field (the sky intensity may vary over the field of view) and the optomechanical constraints of the slit assembly producing discontinuities in the PSF on the detector (significant defocusing at both ends of the sub-slits). The second problem is more critical in the IFU/Argus modes than in the MEDUSA mode because the fibers are smaller, so the contribution of the spectrograph PSF to the overall PSF is larger.

In order to solve the first problem, each sky spectrum is divided by its integrated intensity, so that the interpolation does make sense and provides normalized sky spectra. A model of the sky brightness distribution on the field of view is fitted to the integrated intensities, allowing to recover the intensities of the interpolated sky spectra.

The second problem is solved by adopting the width of the PSF as the free variable in the interpolation process, instead of the y coordinate. The width of the PSF all over the detector should be well known from FF exposures taken through all fibers.

Such a careful sky estimate should allow an efficient subtraction of even strong emission lines, though the energy distribution of the sky is assumed constant all over the field of view, only the overall intensity being allowed to vary. In any case, however, one should keep in mind that at the wavelengths of intense sky emission lines, even a perfect sky subtraction will result in a decrease of S/N.

5. DIRTY FAST PROTOTYPING

While it is quite easy to assess at the early stage of the instrument design the specifications for most technical parameters such as spectral resolution, detector characteristics or thermal stability, the astronomical parameters are much more difficult to guarantee. For GIRAFFE, radial velocity and sky subtraction accuracies are given in the technical specifications[2] and it rapidly appeared that at least a data simulator and some sort of data reduction software are necessary from the beginning of the project to insure that the requirement could be met.

We choose to use the MATLAB to build an ad-hoc simulator and the prototype DRS because of its capability to handle structured data, good experience with the reliability of the programming language and the fact that it was available under university site license on all computers.

We had very rapidly a running software that does not cover all functions of the DRS but permits to verify key performances. For example, the figure 1 shows how the error in the localization of the spectra influences the Radial Velocity (RV) accuracy in the MEDUSA mode. For the high S/N spectra (low RV error) a *plateau* of approximately 2 pixels appears where the error of localization has no effect on the RV. This is not a sort of absolute result and the behavior depends on how the signal is extracted, how the PSF varies, on the amount of diffused light, on the crosstalk between spectra and other physical parameters of the facility. By no means the complete space of possible parameters could be explored, but the possibility to rapidly answer the questions as they appear is crucial.

Two interesting by-products are worth being mentioned. The first by-product is the existence at a very early stage of the project of the working prototype which is manipulated by both astronomer and software specialist. Thus the practical vision of what the software should do and how it is behaving is progressively emerging before the detailed design of the software is freezed and before the first line of the actual code is written. The second by-product is inherent to the used tool. Since the MATLAB scripts are interpreted, the software is dynamically scalable and the full cycle of simulation/reduction could be executed in a matter of seconds interactively if the full scale testing is not necessary. All functionality tests and most accuracy experiments could be done without resort to the full scale simulation and data reduction.

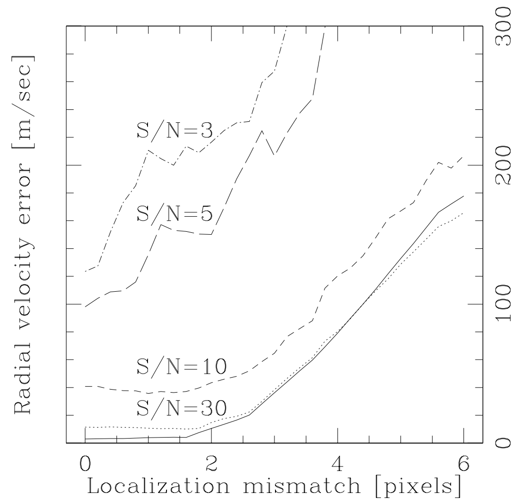

Figure 1. Influence of the localization mismatch on the accuracy of the Radial Velocity. The RV error is increasing with decreasing S/N; curves with S/N=100 (continuous line), 30, 10, 5, 3 are shown.

6. SYSTEM IMPLEMENTATION

Though the easy integration into the fully automatic VLT pipeline (DFS) is the highest priority of the DRS, it could not be completely separated from the interactive off-line usage. The software package such as DRS for any of the VLT instruments relies on the following elements:

1. The modules fulfilling actual data reduction functions
2. The environment
3. The services and utilities available in this environment
4. The data structures

In our case, for the on-line version, the items 2-3 are supplied by the VLT DFS yet to be stabilized and validated (integration is under the responsibility of ESO) and the consensus emerges as to the point 4 where FITS and ASCII is accepted as by-product.

In the off-line version, the items 2-3 are obviously linked and a sensible compromise should be found between the simplicity and the universality of the environment and the volume of the modules to be developed under item 1. For many years astronomers used specialized, observatories-developed environments because of the lack of existing suitable and financially acceptable software. This is the reason of the present situation where several environments compete and every new application written for IRAF, MIDAS, AIPS or any other have the choice to either not use the most interesting specific features available within an environment or to rewrite a substantial part of the package when porting it to another environment. The incompatibility of various environments merely reflects the absence of the standard interfaces similar to higher levels of the X11 standard on which a widely accepted agreement was never reached. It is interesting to notice that there is one exception, the access to data structures where the FITS format resists to 30 years evolution and is largely accepted by the community and even by some commercial packages.

The situation is changing rapidly since the UNIX environment with *gnu* packages is becoming rich enough to provide most of functionalities traditionally supplied by the data analysis system. This leads to the development and availability of utilities depending only on UNIX itself. Though in the VLT the standard environment, beside the UNIX/Posix platform, also the MIDAS environment is available, we decided to develop the software independently of MIDAS (which does not prevent the software to be used from MIDAS) and rely only on UNIX environment and utilities available under UNIX public domain software. This was done after careful evaluation of necessary functions (Functional Specifications) including the graphics and image displays. The link between the DRS and any specific data analysis system, if required, will be through FITS and ASCII data structures only.

6.1. Conclusion

The Data Reduction Software for GIRAFFE spectrograph relies on the specific GIRAFFE feature - the presence of five simultaneous calibration spectra - and the high instrument stability. The localization is adjusted for all exposures and several options are offered for the extraction. The sky subtraction is made via sky intensity distribution and extracted PSF modeling.

The number of software modules is kept low through the use of same modules for multi-objet (MEDUSA) and integral-field spectroscopy (ARGUS and IFU) during most of reduction steps. This is achieved through a careful parameterization of the reduction functions. Associated to the processed image, the numerical mask of *bad pixels* is initialized and progressively updated through the reduction process. The analysis software supports *flagged* pixels all through the analysis. The possibility to replace flagged-pixel at any stage of the data reduction is implemented.

The developed software relies only on UNIX environment and utilities available under UNIX public domain software and standard VLT tools. The link between the DRS and any specific data analysis system, if required, will be through FITS and ASCII data structures only.

6.2. Acknowledgments

This work is supported by the Swiss national fund for scientific research (FNRS). We would like also to thank P. Bartholdi for the helpful discussions and A. Dieny for the work on calibration spectra.

APPENDIX A. PRINCIPAL FUNCTIONS OF THE GIRAFFE DRS

Table 1. Principal functions of the GIRAFFE DRS

name	title
giArithm	Basic operations on numbers, vectors and 2D and 3D images
giGetRemoveBias	Get and remove bias
giSubtractDark	Subtract dark
giEstimatePix	Estimate pixels values in an image
giDetectCosmicS	Detect Cosmic ray hits in a single image
giDetectCosmicM	Detect Cosmic ray hits in several images
giReplaceFlagPix	Replace the flagged pixels
giFitSL	Fit the model to scattered light data
giLocalSpectra	Spectra Localization
giExtractSpectra	Spectra Extraction
giFlatSpectra	Global FF correction - correction for the blaze function, individual fiber transmission and CCD spectral response
giGetWaveSolution	Wavelength Solution
giRebinSpectra	Rebinning in wavelength space
giNormalizeSky	Normalize the Sky spectra
giModelSky	Model the Sky spectrum
giSubtractSky	Subtract Sky spectrum
giFluxCal	Flux spectrophotometric calibration

REFERENCES

1. L. Pasquini, *FLAMES a multiobject fiber facility for VLT*, in SPIE's International Symposium on Astronomical Telescopes and Instrumentation 2000, 2000.
2. GIRAFFE Technical Specifications, ESO Report INS-SPE-ESO-13730-1657,1.0, 1998
3. Functional Specification for BLDR Software, ESO Report, VLT-SPE-OGL-13730-0030, 1999
4. Horne K., 1986, PASP 98, 609
5. M. Viton, B. Milliard, private communication, 1999