DATA REDUCTION PIPELINE FOR EMIR, THE GTC NEAR-IR SPECTROGRAPH

J. Gallego[1], A. Serrano[1], N. Cardiel[1], J. Zamorano[1], J. Gorgas[1], C.E. García-Dabó[1], and A. Gil de Paz[2,3]

[1]Dpto. de Astrofísica, Facultad de Físicas, Universidad Complutense de Madrid, E-28040, Madrid, Spain
[2]NASA/IPAC Extragalactic Database, Caltech, Mail Code 100-22. 770 S. Wilson Ave. Pasadena, CA 91125, USA
[3]Jet Propulsion Laboratory, MS 183-900, 4800 Oak Grove Dr., Pasadena, CA 91109, USA

Abstract

EMIR (Espectrógrafo Multiobjeto Infrarrojo) is a near-infrared wide-field camera and multi-object spectrograph to be built for the 10.4 m GTC (Gran Telescopio Canarias) at La Palma. The Data Reduction Pipeline (DRP), which is being designed by the EMIR Universidad Complutense de Madrid Group, will be optimized for handling and reducing the near-IR data acquired with EMIR. Both reduced data and associated error frames will be delivered to the end-users as a final product.

Key words: ground-based instruments, spectrographs, near-infrared, instrumentation for large telescopes

1. Introduction

GTC (Gran Telescopio Canarias, `www.gtc.iac.es/`) is a 10.4 m telescope to be located at La Palma, Spain. A unique near-infrared wide-field camera and multi-slit spectrograph has been proposed for GTC: EMIR, "Espectrógrafo Multiobjeto Infrarrojo". This will be a state-of-the-art instrument with which multi-object observations will be possible, up to 45 simultaneous spectra with a resolution about 4000 and a spectral coverage from 0.9 to 2.5 microns, over a field of view (FOV) of $6' \times 3'$. It will also have imaging capabilities in the J, H and K near-IR filters. In this case, the FOV is $6' \times 6'$, with a spatial sampling around 0.175 $''$/pixel. The main goal to be achieved is the extension of the study of distant galaxies up to the $z = 2 - 3$ regime, where the rest-frame visible spectrum features are shifted to the K-band. EMIR will allow the analysis of other kind of objects, such as dust-enshrouded star formation regions, low-mass stars and distant clusters of galaxies, among others (Balcells et al. 2000).

EMIR is being developed by a consortium of Spanish, British and French institutions, led by the Instituto de Astrofísica de Canarias (IAC), and which also includes the Universidad Complutense de Madrid (UCM), the Instituto Nacional de Técnica Aeroespacial (INTA), the University of Durham and the Observatoire Midi-Pyrénées (Toulouse). For further details, visit the EMIR web at `www.ucm.es/info/emir/`.

Due to the complexity of EMIR, a great amount of data has to be analyzed. The Data Reduction Pipeline (DRP hereafter), which is being designed by the UCM Group, is intended to reduce the data acquired with EMIR as a part of the GTC *Control System*. It will use a set of robust and fail-safe software tools and algorithms specialized in near-IR observations.

In Section 2, an overview of the DRP, as well as the main goals which should be achieved, are presented. In Section 3, we give a description of the data flow within the DRP. In Section 4, the system key features are listed and, finally, in Section 5, immediate future work is briefly summarized.

2. Overview of the DRP

The DRP will constitute a set of advanced reduction algorithms and the quality check rules to be followed. It will be integrated into the GTC *Data Factory* and accessed through the GTC *Inspector*. This will create a common framework in which EMIR, and other instruments specialized software, could be executed. For each of the EMIR observing modes, imaging and multi-object spectroscopy, there are associated data types, including calibration observations (darks, flats, etc.) and science frames. Once an observation is completed, the DRP will need to interact with the GTC *Operation Repository*, where the existing calibration frames and the results are stored. Then, the DRP will provide the users with a final reduction set of observations in physical units and its associated error frames.

The DRP will be coded under an object-oriented architecture (C++), following GTC programming and software standards. This will minimize porting problems to different operating systems and platforms. It will work under Sun Solaris, but if necessary, it will also be tested for VxWorks, which are the GTC hardware standards. Currently, the DRP is not intended to be public, though this option is being taken into consideration. It is being developed under the Rational Unified Process (RUP) methodology (Kruchten 1999) and Unified Modeling Language (UML) symbology.

The DRP has to tackle several critical problems to handle and reduce the near-IR data obtained by EMIR:

1. The suppression of OH sky lines is a key problem. These spectral features are due to hydroxyl radicals, which produce a dense *forest* of emission lines, especially in the $1 - 2.5$ microns region (Rousselot et al.

Proc. Symposium *'The Promise of the Herschel Space Observatory'* 12–15 December 2000, Toledo, Spain
ESA SP-460, July 2001, eds. G.L. Pilbratt, J. Cernicharo, A.M. Heras, T. Prusti, & R. Harris

2000). Current near-IR spectrographs in the J and H bands use high spectral resolution ($R > 2000$) and mask the pixels where the OH lines are detected. The goal of EMIR is to apply the same technique to the three J, H and K bands.
2. In the near-IR, the detection of astronomical sources is a very difficult issue, as they are much fainter than the sky background. To solve this problem, a large number of short exposures must be co-added.
3. Up to 45 simultaneous spectra can be achieved in the multi-object spectroscopic observing mode. Such quantity of data has to be handled properly.
4. A quick way to visualize the data being gathered must be provided.
5. The user will need a reliable error estimation. Errors associated to each image shall be carefully tracked.

3. Data Flow of the DRP

Raw data in the detector are actually a convolution of the source astronomical signal plus the different contributions from each of the media the light travels through, i.e., the atmosphere, the telescope, the instrument and, finally, the detector itself. To uncover the scientific information, a data reduction must be performed. The recipes to be followed are different for each type of data, i.e., calibration or science frames, and also depend on the observing mode (imaging or spectroscopic). The Reduction Process or Data Flow to be followed in each situation, graphically summarized in Figure 1, is described in the following subsections.

3.1. Raw Calibration Frame Process

Calibration frames can be darks, flatfields, sky frames or line lamp frames. Both in imaging and spectroscopic observing modes, they need to be cleaned with operations including:

1. Cosmic rays removal, with a two-dimensional interpolation between adjacent pixels.
2. Cosmetic masking for defects in the detector, such as dead pixels, also with a two-dimensional interpolation.
3. Single frame averages, for combinations on a pixel-to-pixel basis.

Each calibration frame needs further specific reduction recipes. In concrete:

- Darks are only averaged after being cleaned. They are used for dark subtraction, which is important even for near-IR cryostatically cooled instruments like EMIR.
- Flatfields need a dark subtraction. Then they have to be averaged and normalized. They are used to correct for pixel response non-uniformity in the detector.
- Sky frames need dark subtraction and averaging. They are important because of the high and time-changing contribution of the sky background in the near-IR. They can be achieved by beam switching or dithering techniques in imaging mode. Beam switching can also be used in multi-slit spectroscopic mode, as well as nodding.
- Line lamp frames are just averaged and do not need any dark subtraction.

Once the raw calibration frame has been reduced, this calibration image is ready to be used in the reduction of raw science frames.

3.2. Raw Science Frame Process

Science frames, which correspond to astronomical objects of interest, suffer a longer process until the final reduction, which can be divided into two phases: Pre-Process and Post-Process.

3.2.1. Science Frame Pre-Process

In this phase the calibration frames described in Section 3.1 are used, as raw science frames must be corrected with sky and dark subtraction, flatfielding, cosmic rays and cosmetic defects removal.

3.2.2. Science Frame Post-Process

This depends on the observing mode:

- *Imaging Mode:*
  In this case, there are three important operations:
  1. Image restoration, where geometric distortions which have not been removed during pre-processing, have to be eliminated. This includes dithering techniques, when offsets between science and sky exposures are small compared with the field size, and mosaicking, when the offsets are larger.
  2. Atmospheric extinction correction.
  3. Absolute Flux calibration.
- *Multi-slit Spectroscopic Mode:*
  1. Wavelength calibration using line lamp frames.
  2. One-dimensional spectra identification and extraction from the multi-object spectroscopic frames, where there can be up to 45 simultaneous sources spectra. These two-dimensional frames are analyzed, so that each spectrum is registered and recorded as a one-dimensional frame.
  3. OH-lines masking, in order to correct saturated pixels by the intense emission of hydroxyl radicals.
  4. Image restoration, to correct curved slit images (C-distortion) and/or curved spectra in the wavelength direction (S-distortion).
  5. Atmospheric extinction correction.
  6. Flux calibration.

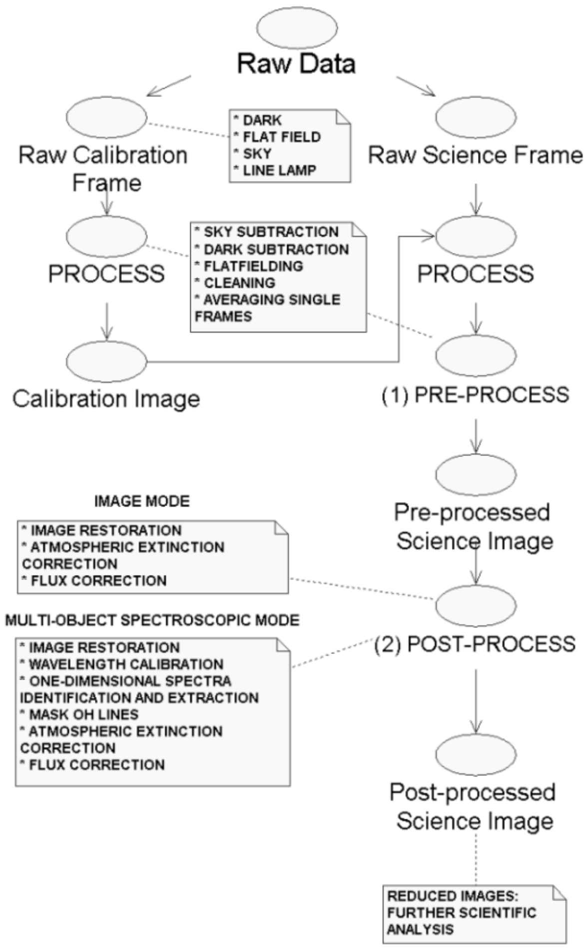

Figure 1. Data Flow of the DRP.

4. Key Features of the DRP

Next items give an overview of the most important features that the DRP should include.

4.1. Object-oriented architecture

The whole DRP, which is embedded into the GTC *Data Factory*, will be coded under an object-oriented architecture. This point of view allows to interpret an observation as an *object*, i.e., an *instance* of a C++ class created for observations. This class will include the image matrix, additional parameters including calibrations applied and the matrix for the associated errors. The *methods* of this class (called *member functions* in C++) will be the different reduction operations to be performed in this image.

4.2. FITS as default delivering format

FITS will be the default format for delivering the results to the scientific community. As an added feature, data will be portable to the most common format used by the astronomical community at the time the instrument is operating.

4.3. Quick-look facility

This component is intended to allow a fast examination of raw frames and pre-processed images. It will be used to visualize sets of data and to monitor the observation in real-time. It will offer a wide variety of graphical resources, as well as preliminary inspection tools specific to the observing mode, such as simple statistics, zooms, cuts and radial profiles, among others.

4.4. Error propagation

The DRP will track error propagation via error images associated to each data image, and generated just after reading out the raw frames, including the readout noise and gain. A given noise model has to be taken into account. Error frames will be processed in parallel to data frames.

4.5. Image restoration

The DRP will be able to automatically remove field distortions in images.

4.6. History of each image

All the operations performed to any image will be recorded in the image header and/or in a general log file. With this history file, a reduction process could be undone to recover the original frames.

4.7. Fully automatic reduction mode

The DRP will offer a fully automatic running mode, where standard assumptions based on previous experience will be assumed to perform a silent reduction. All the relevant information will be always recorded in the history of the image file.

4.8. Step-by-step interactive reduction mode

In this interactive mode the user will control, either in a graphical or command-driven way, the evolution of the reduction procedure. It will be possible to use a running mode with some steps in an automatic mode and others, especially those considered, in an interactive mode.

4.9. Stand-alone version

The DRP is intended to be fully integrated into the GTC *Control System*, in particular, into the GTC *Data Factory*. This means that this software will not be of public access, but will be used at the GTC facilities. However, a stand-alone version to execute a processing recipe independently of GTC *Control Software* is currently under consideration.

4.10. Standard reduction recipes

As part of the DRP, the user will be provided with a set of standard reduction recipes for each astronomical problem. They will include a suggested data flow and detailed advice about how to perform it from the beginning.

4.11. Checks and quality control procedures

Checks and quality control procedures will be also included to provide the user with tools to estimate the quality of the data.

4.12. Solutions specific to each observing mode

The DRP will treat each observing mode as an independent problem. In this way, optimized solutions will be implemented when necessary. However, whenever possible, already developed software will be reused for different EMIR modes.

4.13. Quick astrometry and object identification

The DRP will be able to produce a preliminary catalogue of sources present in the frame to perform an astrometric solution (about $2''$ accuracy) and to check which of the entries in the catalogue were previously present in a predefined database.

4.14. Graphic and command-line interfaces

A user-friendly graphic user interface will be available as default through the GTC *Inspector*. The user will have also available a command driven interface. The DRP should allow user-friendly scripting.

4.15. Guaranteed for Sun Solaris

The DRP software shall be guaranteed and tested for Sun Solaris, following the GTC hardware standards.

4.16. User Manual

The DRP User Manual will describe the system from both the astronomer and the engineer point of view. The following characteristics would be important:

1. Different levels of complexity.
2. Guidelines for every main reduction step.
3. Detailed general reduction recipes.
4. Known bugs and workarounds.
5. Available at the most common printing formats.
6. A Frequently Asked Questions (FAQ) section.

4.17. On-line help and documentation

The DRP shall provide an on-line help system to assist the user. Each topic covered in the User Manual will also be available through the on-line help.

5. Future work

GTC Telescope is intended to start scientific operations in 2003. However, EMIR itself is not expected to be fully operational until mid-2005. The EMIR DRP has recently passed an inception phase, in which the most important and useful functionalities and requirements have been identified. From now on, the UCM Group will be working at an elaboration phase, in which the overall DRP architecture is to be defined and depicted. Also a near complete description of the system has to be developed. A simplified software prototype is intended to be coded in order to find and solve critical risks as soon as possible.

Acknowledgements

EMIR is being financed by the GTC instrumentation program and the EMIR partner institutions, which are, the Instituto de Astrofísica de Canarias (IAC), the Universidad Complutense de Madrid (UCM), the Instituto Nacional de Técnica Aeroespacial (INTA), the University of Durham and the Observatoire Midi-Pyrénées (Toulouse).

References

Balcells, M., Guzmán, R., Patrón, J., et al, 2000, SPIE AS06 conf. 4008-86-2000, Munich

Kruchten, P., 1999, *The Rational Unified Process: an Introduction*, Addison Wesley Longman

Rousselot, P., Lidman, C., Cuby, J.-G., Moreels, G., Monnet, G., 2000, A&A, 354, 1134