Data reduction pipeline for EMIR: a near-IR multi-object spectrograph for the Spainish 10-m telescope

Jesus Gallego, Nicolas Cardiel, Angel Serrano, Javier Gorgas, Jaime Zamorano, et al.

Jesus Gallego, Nicolas Cardiel, Angel Serrano, Javier Gorgas, Jaime Zamorano, Francisco Garzon, Cesar-Enrique Garcia-Dabo, Armando Gil de Paz, "Data reduction pipeline for EMIR: a near-IR multi-object spectrograph for the Spainish 10-m telescope," Proc. SPIE 4847, Astronomical Data Analysis II, (19 December 2002); doi: 10.1117/12.460344

Event: Astronomical Telescopes and Instrumentation, 2002, Waikoloa, Hawai'i, United States

Data Reduction Pipeline for EMIR, a near-IR multiobject spectrograph for the Spanish 10m telescope

Jesús Gallego[a], Nicolás Cardiel[a], Ángel Serrano[a], Javier Gorgas[a], Jaime Zamorano[a], F. Garzón[b], César-Enrique García-Dabó and Armando Gil de Paz[c,d]

[a]Departamento de Astrofísica y Ciencias de la Atmósfera, Universidad Complutense de Madrid, E-28040, Madrid, Spain

[b]Instituto de Astrofísica de Canarias, E-28200, Tenerife, Spain

[c]NASA/IPAC Extragalactic Database, Caltech, Pasadena, CA 91125, USA

[d]Jet Propulsion Laboratory, 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 10.4m Spanish telescope (Gran Telescopio Canarias, GTC) at La Palma. The Data Reduction Pipeline (DRP), which is being designed and built by the EMIR Universidad Complutense de Madrid group, will be optimized for handling and reducing near-infrared data acquired with EMIR. Both reduced data and associated error frames will be delivered to the end-users as a final product.

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

1. INTRODUCTION

Gran Telescopio Canarias (hereafter GTC, `http://www.gtc.iac.es/`) is a 10.4m telescope to be located at the Roque de los Muchachos Astronomical Observatory 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 spectroscopic observations will be possible for up to 45 simultaneous targets with a resolution about 4000 and a spectral coverage from 0.9 to 2.5 microns. The field of view (FOV) will be $6' \times 3'$ in spectroscopic mode. EMIR will also have imaging capabilities in the J, H and K near-IR bands. In this case, the FOV is $6' \times 6'$, with a spatial sampling around 0.175 "/pixel.

The main scientific driver of EMIR is the extension of the study of distant galaxies up to the $z = 2 - 3$ redshift regime, where the rest-frame visible spectrum features are shifted to the K-band. EMIR will also 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 (see[1] for a more detailed description).

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

Due to the nature and complexity of EMIR, it will produce a great amount of data that will need to be analyzed. The EMIR Data Reduction Pipeline (hereafter DRP), which is being developed by the EMIR 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 data.

Further author information: Send correspondence to J. Gallego: E-mail: jgm@astrax.fis.ucm.es

Astronomical Data Analysis II, Jean-Luc Starck, Fionn D. Murtagh, Editors, Proceedings of SPIE Vol. 4847 (2002) © 2002 SPIE · 0277-786X/02/$15.00

In Section 2, an overview of the DRP, as well as the main goals which shall be achieved, are presented. In Section 3, we provide 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 be 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 ([2]), 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 molecule, which produce a dense *forest* of emission lines, especially in the $1 - 2.5$ microns region. Current near-IR spectrographs in the J and H bands use high spectral resolution ($R > 4000$) 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 could be much fainter than the sky background. To solve this problem, a large number of short exposures must be co-added.
3. In the near-IR, the sky can change in a scale of minutes. To have this effect into account, the process of estimating the appropriate sky background for a given image is a difficult task that has to be undertaken carefully. Moreover, the background is contaminated by a large number of astronomical sources that have to be removed when estimating the sky.
4. Up to 45 simultaneous spectra can be achieved in the multi-object spectroscopic observing mode. Such quantity of data has to be handled properly.
5. Given the complex optical design of the instrument, geometric distortions are expected to be severe. They will need to be taken into account.
6. The user will need a reliable error estimation. Errors associated to each image shall be carefully tracked.
7. A quick way to visualize the data being gathered must be provided.

3. DATA FLOW OF THE DRP

Raw data in the detector is 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 contained in this noisy signal, a data reduction must be performed.

This will depend on the type of data, those related to calibration purposes, such as darks and flatfields, and those of science interest, as well as on the observing mode (imaging or spectroscopic). The Reduction Process or Data Flow followed in the different cases can be graphically summarized in Figure 1, and will be explained in the following subsections.

3.1. Raw Calibration Frame Process

Calibration Frames can be darks, flatfields, sky frames or line lamp frames. The process needed to reduce them, both in imaging and spectroscopic observing modes, makes use of several of the following five operations, depending on the type of frame:

1. Dark subtraction, which is important even in the near-IR cryostatically cooled instruments like EMIR.
2. Flatfield division, in order to correct for pixel response non-uniformity in the detector.
3. Cosmetic masking for defects in the detector, such as dead pixels, also with a two-dimensional interpolation.
4. Cosmic rays removal, with a two-dimensional interpolation between neighbor adjacent pixels.
5. Sky subtraction, where the high and time-changing contribution of the sky background is eliminated from raw frames. This fundamental step can be taken into account by beam switching or dithering techniques in imaging mode. Beam switching can also be used in multi-slit spectroscopic mode, as well as nodding.

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 total reduction, which can be divided into two clear phases: Pre-Process and Post-Process.

3.2.1. Science Frame Pre-Process

This phase is similar to that reported at 3.1, where raw science frames must be corrected with sky and dark subtraction, flatfielding, cosmetic defects and cosmic rays removal. To do so, several calibration images need to be used.

3.2.2. Science Frame Post-Process

After being pre-processed, a science frame needs further recipes to completely end its reduction. 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 concerns to dithering techniques, where offsets between science and sky exposures are small compared with the field size, and also when mosaicking, where the offsets are larger.
2. Atmospheric extinction correction, due to the changing absorption of light with different airmasses and wavelengths.
3. Flux calibration, in order to measure in absolute units the incident flux out of the atmosphere from the observed objects.

Multi-slit Spectroscopic Mode:

As well as the three operations in the previous section, the spectroscopic mode is characterized by:

1. Wavelength calibration using line lamp frames, since the available dispersion elements do not provide spectra with a perfect linear dispersion.
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 high intense emission of hydroxyl molecule.

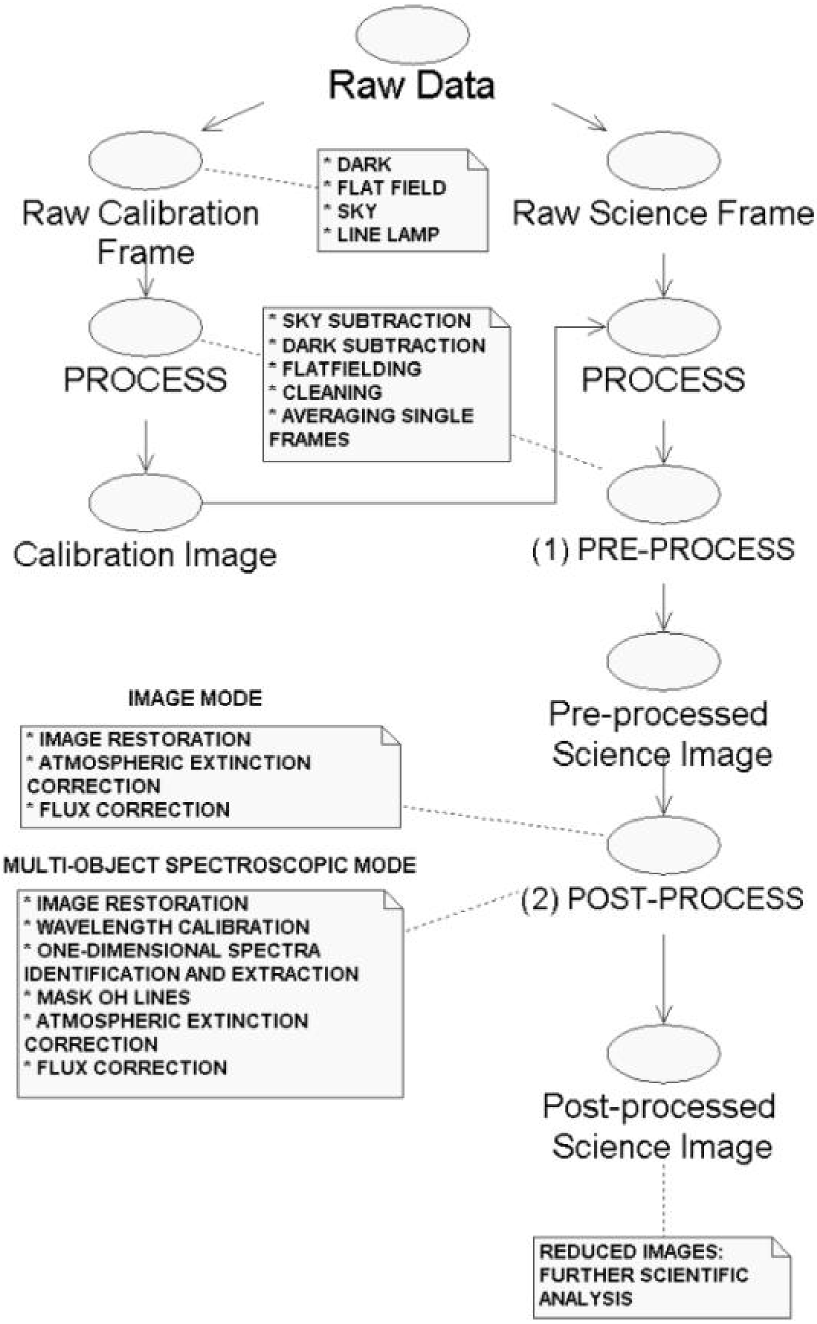

Figure 1. Data Flow of the DRP.

4. KEY FEATURES OF THE DRP

In the following lines, there is an overview of the most important features that the EMIR DRP will include.

4.1. Object-oriented architecture

The whole DRP, which is embedded into the GTC Data Factory, shall be coded under an object-oriented architecture. This point of view allows to interpret an observation as an "object", this is, an instance of a C++ class created for observations. This class will include the image (as a matrix), additional parameters including calibrations applied, and another 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, so both raw and final data shall be FITS files. As an added feature, data shall be portable to the most common format used by the astronomical community when 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 monitor in real-time the observation. It shall offer a wide variety of graphical resources, as well as preliminary inspection tools specific to the observing mode, such as simple statistics, zooms, cuts, radial profiles, among others.

4.4. Error propagation

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

4.5. Image restoration

The DRP shall be able to automatically remove field distortions in images when mosaicking and dithering.

4.6. History of each image

All the operations performed to any image shall be recorded, for example, in the image header and/or in a general log file.

4.7. Fully automatic reduction mode

The DRP shall offer a fully automatic running mode, where standard assumptions based on previous experience shall be assumed to perform a silent reduction. All the relevant information shall 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 by graphical or command driven way the evolution of the reduction procedure. It shall be possible to use a running mode with steps in an automatic mode and steps performed in an interactive mode, especially those considered critical.

4.9. Stand alone version

The DRP is intended to be fully integrated into the GTC Control System, in special, into the GTC Data Factory. This means that this software will not be of public access, but will be used at GTC facilities. However, a stand alone version which allows 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 shall be provided with a set of standard reduction recipes for each astronomical problem. They shall include a suggested data flow and detailed advice about how to perform it from the beginning.

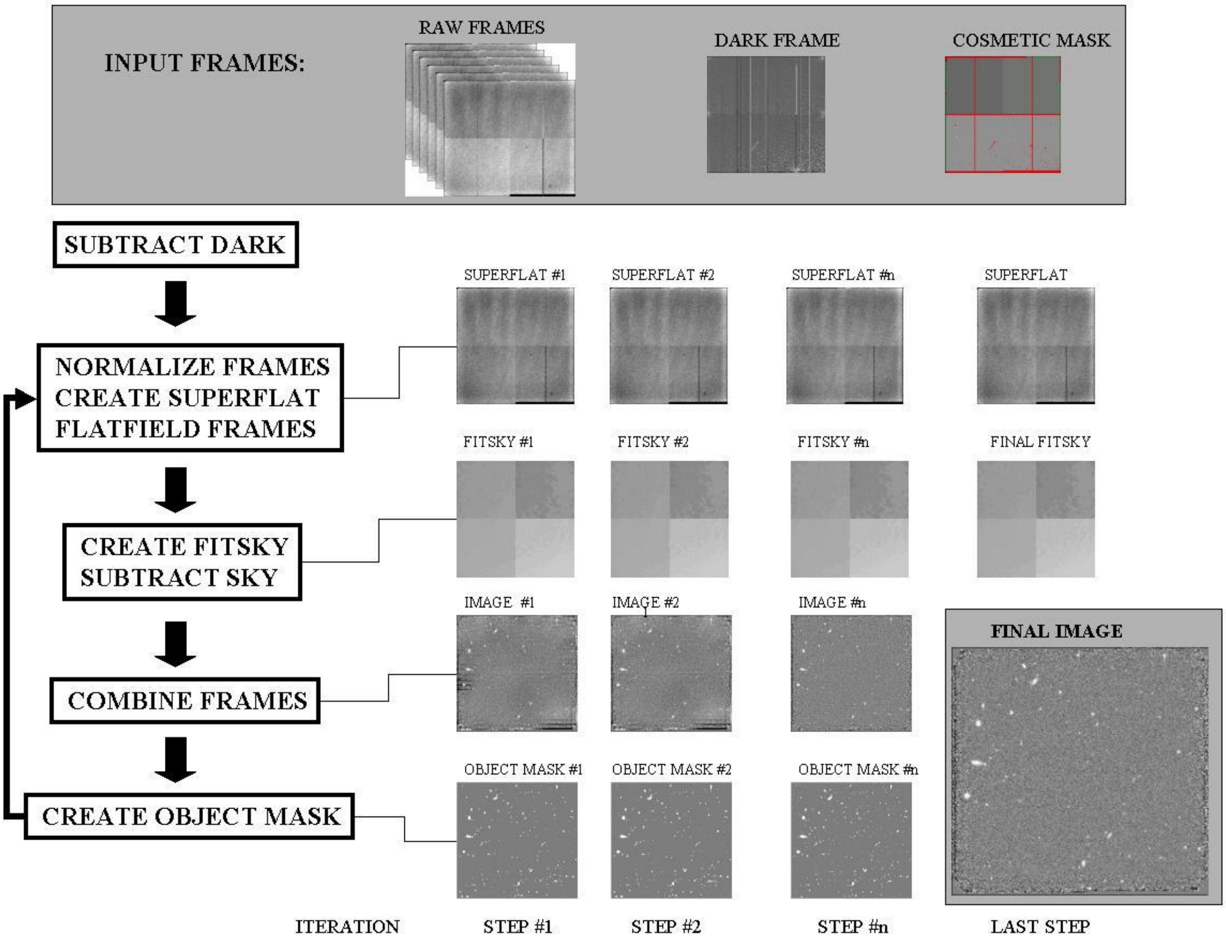

Figure 2. Example of reduction procedure for direct images taken with EMIR.

4.11. Checks and quality control procedures

Checks and quality control procedures shall 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 tackle each observing mode as an independent problem. In this sense, optimized solutions shall be implemented when necessary. However, whenever possible, already developed software shall be reused for different EMIR modes.

4.13. Quick astrometry and object identification

The DRP shall be able to produce a preliminary catalog of sources present in the frame, perform an astrometric solution (about 2" accuracy) and check which of the entries in the catalog were previously present in a predefined database. The whole process shall consume no less than two minutes.

4.14. Graphic and command-line interfaces

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

The GTC standard for software documentation is the Doxygen (`http://www.doxygen.org/`) documentation system. This is the format that has been adopted for the DRP.

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 shall also be available through the on-line help.

5. FUTURE WORK

GTC Telescope is intended to start scientific operations in 2004. However, EMIR itself is expected to be fully operational in the 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, our UCM Group will be working at an elaboration phase, in which the overall DRP architecture is to be defined and depicted, as well as a near complete description of the system has to be developed. A simplified software prototype is has been coded and is being tested, in order to find and solve critical risks as soon as possible.

ACKNOWLEDGMENTS

EMIR is being financed by the GTC instrumentation program and the EMIR partner institutions, which are, the Instituto de Astrofísica de Canarias (IAC, Spain), the Universidad Complutense de Madrid (UCM, Spain), the University of Durham (UK) and the Observatoire Midi-Pyrénees (Toulouse, France).

REFERENCES

1. M. Balcells, R. Guzman, J. Patron, A. Aragon-Salamanca, J. Azcue, J. A. Ballester Lluch, M. T. Barroso, F. Beigbeder, S. Brau-Nogue, N. Cardiel, D. Carter, J. J. Diaz-Garcia, E. de la Fuente, F. J. Fuentes, A. B. Fragoso-Lopez, F. Gago, J. Gallego, J. Gomez-Elvira, J. C. Heredero, D. J. Jones, J. C. Lopez, P. Luke, A. Manescau, T. Munoz, R. F. Peletier, R. Pello, J. P. Picat, D. J. Robertson, J. A. Rodriguez, A. Serrano, R. M. Sharples, and J. Zamorano, "EMIR: cryogenic NIR multi-object spectrograph for GTC," in *Proc. SPIE Vol. 4008, p. 797-805, Optical and IR Telescope Instrumentation and Detectors, Masanori Iye; Alan F. Moorwood; Eds.*, **4008**, pp. 797–805, Aug. 2000.
2. P. Kruchten, *The Rational Unified Process: an introduction*, Addison Wesley Longman, Massachusets, 1999.

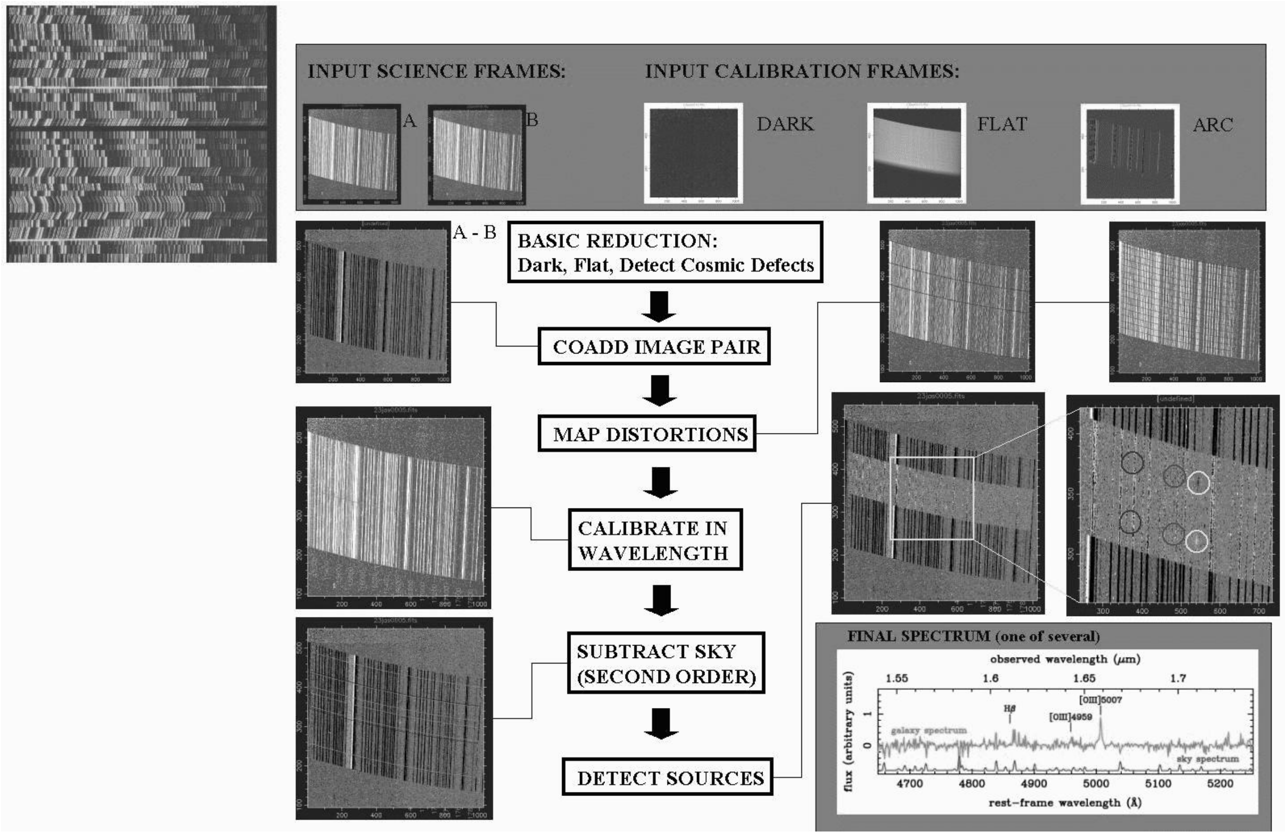

Figure 3. Example of reduction procedure for images taken with EMIR in spectroscopic mode.