RevMexAA (Serie de Conferencias), **16**, 73–76 (2003)

A NEW APPROACH IN DATA REDUCTION: PROPER HANDLING OF RANDOM ERRORS AND IMAGE DISTORTIONS

N. Cardiel, J. Gorgas, J. Gallego, Á. Serrano, and J. Zamorano

Departamento de Astrofísica, Universidad Complutense de Madrid, Spain

M. L. García-Vargas, P. Gómez-Cambronero, and J. M. Filgueira

Gran Telescopio Canarias, La Laguna, Spain

RESUMEN

Los procesos de reducción de datos tienen como objetivo minimizar el impacto que las imperfecciones en la adquisición de los mismos producen en la obtención de medidas de interés para el astrónomo. Para conseguir este objetivo, es necesario realizar manipulaciones aritméticas, utilizando imágenes de datos y de calibración. Por otro lado, la interpretación correcta de las medidas sólo es posible cuando existe una determinación precisa de los errores asociados. En este trabajo discutimos diferentes estrategias posibles para obtener determinaciones realistas de los errores aleatorios finales. En concreto, destacamos los beneficios que conlleva considerar el proceso de reducción de datos como la caracterización completa de las imágenes originales, pero evitando, tanto como sea posible, la alteración aritmética de las imágenes hasta el momento de su análisis final y obtención de medidas definitivas. Esta filosofía de reducción será utilizada en la reducción de datos de ELMER y de EMIR.

ABSTRACT

Data reduction procedures aim to minimize the impact of data acquisition imperfections on the measurement of data properties with a scientific meaning for the astronomer. To achieve this purpose, appropriate arithmetical manipulations with data and calibration frames must be performed. Furthermore, a full understanding of all the possible measurements relies on the firm constraint of their associated errors. We discuss different strategies for obtaining realistic determinations of final random errors. In particular, we highlight the benefits of considering the data reduction process as the full characterization of the raw data frames, but avoiding, as far as possible, the arithmetical manipulation of the data until the final measurement and analysis of the image properties. This approach will be used in the pipeline data reduction for ELMER and EMIR.

Key Words: **METHODS: ANALYTICAL — METHODS: DATA ANALYSIS — METHODS: NUMERICAL — METHODS: STATISTICAL**

Science with the GTC 10-m Telescope (Granada, España, 5-8 febrero 2002)
Editors: José Miguel Rodríguez Espinosa, Francisco Garzón López & Verónica Melo Martín

1. INTRODUCTION

The Gran Telescopio Canarias (GTC)[1], as one the best human tools for exploring and revealing the unknown Universe, will give access, in conjunction with its pioneering instrumentation, to very faint and/or distant objects, in practice inaccesible for 4 m class telescopes. For that reason, very high signal-to-noise ratios are expected to be uncommon in most cases. Under these circumstances, accurate error estimation is essential to guarantee the reliability of the measurements.

Although there are no magical recipes for quantifying systematic errors in a general way, a case-by-case solution needing be sought, the situation is, fortunately, not so bad concerning random errors. Initially, the latter can be measured and properly handled using standard statistical tools. In this contribution, we discuss the benefits and drawbacks of various methods of quantifying random errors in the context of data reduction pipelines. After examining the possibilities, we conclude that the classical reduction procedure is not perfectly suited for error handling. In this sense, the responsibility for the completion of the more complex data reduction steps must be transferred to the analysis tools. For this approach to be possible, additional information must also be provided for those tools, which in turn implies that the reduction process should be modified in order to produce that information. A discussion concerning the treatment of systematic errors is beyond the scope of this paper.

[1] http://www.gtc.iac.es

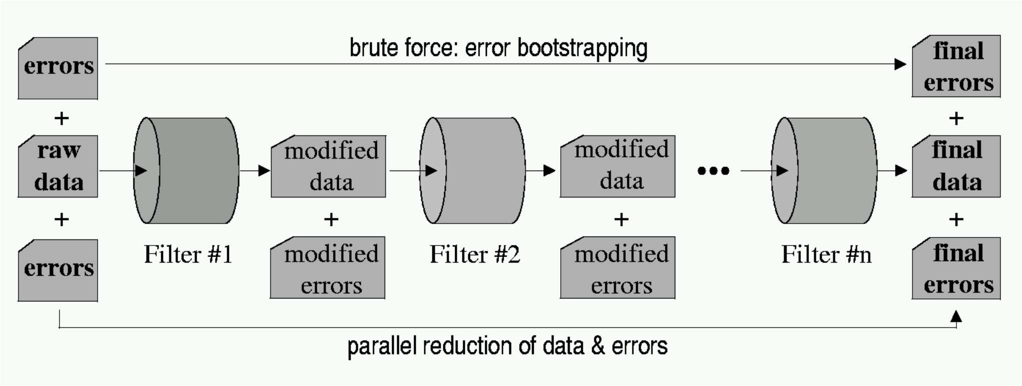

Fig. 1. Classical reduction procedure.

2. THE CLASSIC REDUCTION PROCEDURE

2.1. *Three methods of quantifying random errors*

According to the classic view (see Figure 1), a typical data reduction pipeline can be considered as a collection of filters, each of which transforms input images into new output images after performing some kind of arithmetical manipulation and making use of additional measurements and calibration frames when required. In this scenario, three different approaches may in principle be employed to determine random errors in completely reduced images:

1. *Comparison of independent repeated measurements.* This is one of the simplest and most straightforward ways to estimate errors, since, in practice, errors are not computed nor handled through the reduction procedure. The only requirement is the availability of a non too small number of independent measurements. Although as such can be considered even the flux collected by each *independent* pixel in a detector (for example when determining the sky flux error in direct imaging), in most cases this method requires the comparison of different frames. For that reason, and given that for many purposes it may constitute an extremely expensive method in terms of observing time, its applicability on a general situation seems rather unlikely.

2. *First principles and brute force: error bootstrapping.* Exploiting our knowledge concerning how photo-electrons are generated (expected statistical distribution of photon arrival into each pixel, detector gain and read-out noise), it is possible to generate an error image associated to each raw-data frame. By means of error bootstrapping via Monte Carlo simulations, new instances of the initial raw-data frame are simulated and can be completely reduced as if they were real observations. Comparison of the measurements performed over the whole set of reduced simulated observations provides then a good estimation of the final errors. However, even though this method results in less observing time wasted, it can also be terribly expensive, but now in terms of computing time.

3. *First principles and elegance: parallel reduction of error and data frames.* Instead of wasting either observing or computing time, it is also possible to feed the data reduction pipeline with both the original raw data frame and its associated error frame (calculated from first principles) without repetition of this process throughout the whole reduction process. In this case every single arithmetical manipulation performed over the data image must be translated, using the law of propagation of errors, into parallel manipulations of the error image. Unfortunately, the data reduction packages normally used in astronomy (e.g., IRAF, MIDAS, etc.) do not consider random error propagation as a default operation; thus, some kind of additional programming is unavoidable.

2.2. *Error correlation—a real problem*

Although each of the three methods described above is suitable for use in different circumstances, the third approach is undoubtedly that which, in practice, may be used in a more general situation. In fact, once the appropriate data reduction tool is available, the parallel reduction of data and error frames is the only way to proceed when observing or computing time demands are prohibitively high. However, because of the unavoidable fact that the information collected by detectors is physically sam-

Science with the GTC 10-m Telescope (Granada, España, 5-8 febrero 2002)
Editors: José Miguel Rodríguez Espinosa, Francisco Garzón López & Verónica Melo Martín

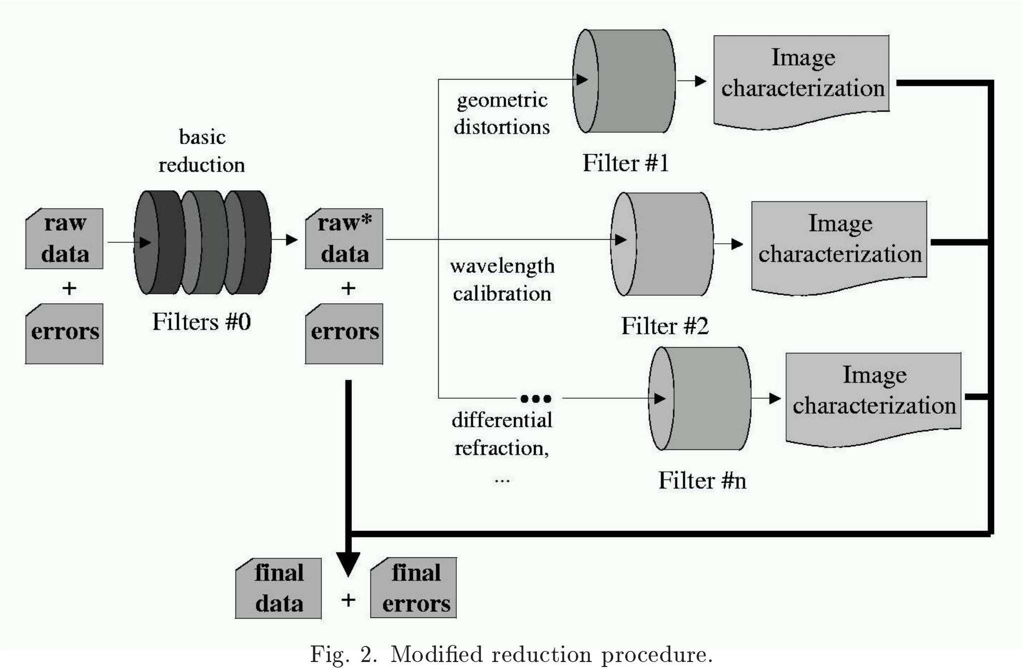

Fig. 2. Modified reduction procedure.

pled in pixels, this approach comes up against a major problem: errors start to become correlated as soon as one introduces image manipulations involving rebinning or non-integer pixel shifts of data. A naive use of analysis tools would neglect the effect of covariance terms, leading to dangerously underestimated final random errors. Actually, this is probably the most common situation since, initially, the classical reduction package operates as a black box, unless specially modified to the contrary. Unfortunately, as soon as one accumulates a few reduction steps involving increments of correlation between adjacent pixels (e.g., image rectification when correcting for geometric distortions, wavelength calibration into a linear scale, etc.), the number of covariance terms starts to increase too rapidly to make feasible the possibility of stacking up and propagating all the new coefficients for every pixel of an image.

3. THE MODIFIED REDUCTION PROCEDURE

3.1. *Image characterization*

Obviously, the emergence of the problem can be prevented, if, for example, one does not allow the data reduction process to introduce correlation into neighboring pixels before the final analysis. In other words, if all the reduction steps that lead to error correlation are performed in a single step during the measurement of the image properties with a scientific meaning for the astronomer, there are no previous covariance terms to be concerned with. Whether this is actually possible or not may depend on the type of reduction steps under consideration. In any case, a change in the philosophy of the classical reduction procedure can greatly help in alleviating the problem. The heart of this change consists in considering the reduction steps that originate pixel correlation as filters that *do not necessarily* take input images and generate new versions of them after applying some kind of arithmetical manipulation, but as filters that properly *characterize* the image properties, without modifying those input images.

More precisely, the reduction steps can be divided into two groups (see Figure 2): a) simple steps, which do not require data rebinning or non-integer pixel shifts of data; and b) complex steps, those suitable for introducing error correlation between adjacent pixels. The former may be operated as in classical reduction, since their application does not introduce covariance terms. However, the complex steps are allowed to determine only the required image properties that one would need to actually perform the correction. For more common situations, these characterizations may be simple polynomials (in order to model geometric distortions, non-linear wavelength calibration scales, differential refraction dependence with wavelength, etc.). According this view, the end product of the modified reduction procedure is constituted by a slightly modified version of the raw data frames (after quite simple arithmetical manipulations) and by an associated collection of image characterizations.

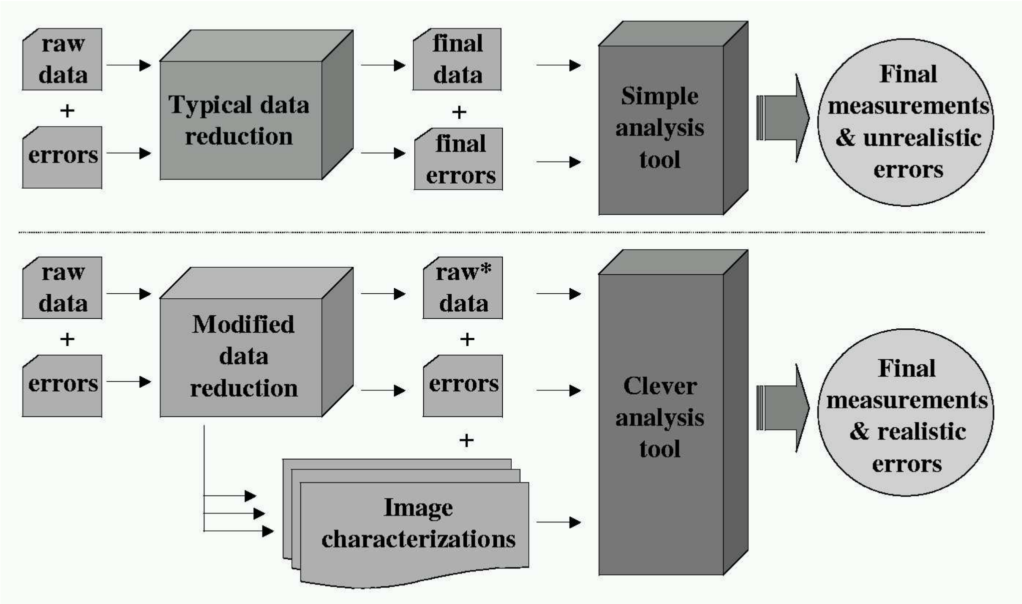

Fig. 3. Comparison between classical (upper panel) and modified (lower panel) reduction procedures.

3.2. *Modus operandi*

Clearly, at any moment it is possible to combine the result of the partial reduction after all the linkable simple steps, with the information achieved through all the characterizations derived from the complex steps, to obtain the same result as in classical data reduction (thick line in Figure 2). However, instead of trying to obtain completely reduced images ready for starting the analysis work, one can directly feed a *clever analysis tool* with the end products of the modified reduction procedure (see Figure 3). Obviously, this clever analysis tool has to perform its task taking into account that some reduction steps have not been performed. For instance, if one considers the study of a 2D spectroscopic image, the analysis tool should use the information concerning geometric distortions, wavelength calibration scale, differential refraction, etc., to obtain, for example,

an equivalent width through the measurement in the partially reduced image (uncorrected for geometric distortions, wavelength calibration, etc.). To accomplish this task, it is necessary to manipulate the data using a new and distorted system of coordinates that must override the orthogonal coordinate system defined by the physical pixels. It is in this step where the final error of the equivalent width should be obtained. It is important to highlight that, in this situation, such error estimation should not be a complex task, since the analysis tool is supposed to be handling uncorrelated pixels.

The described reduction philosophy will be incorporated into the pipeline data reduction for ELMER[2] and EMIR.[3]

Financial support for this research was provided by Spain's Programa Nacional de Astronomía y Astrofísica under the grants AYA2000-977 and AYA-2000-1790. This work has benefited from experience gained in reducing NIRSPEC data obtained at Keck II. In this sense, N. C. acknowledges partial financial support from a UCM Fundación del Amo Fellowship and a short contract at UCSC.

N. Cardiel, J. Gorgas, J. Gallego, Á. Serrano, and J. Zamorano: Departamento de Astrofísica, Facultad de Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain (ncl@astrax.fis.ucm.es)
M. L. García-Vargas, P. Gómez-Cambronero, and J. M. Filgueira: Gran Telescopio Canarias (GTC), C/ Vía Láctea s/n (Instituto de Astrofísica de Canarias), E-38205 La Laguna, Tenerife, Spain

[2]See http://www.gtc.iac.es/instrumentation/elmer_s.asp.
[3]See http://www.ucm.es/info/emir.

Science with the GTC 10-m Telescope (Granada, España, 5-8 febrero 2002)
Editors: José Miguel Rodríguez Espinosa, Francisco Garzón López & Verónica Melo Martín