PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, **116**:148–153, 2004 February

A Fast Algorithm for Cosmic-Ray Removal from Single Images

WOJTEK PYCH

David Dunlap Observatory, University of Toronto, P.O. Box 360, Richmond Hill, ON L4C 4Y6, Canada;
and Copernicus Astronomical Center, Bartycka 18, PL-00-716 Warszawa, Poland; pych@camk.edu.pl

Received 2003 November 12; accepted 2003 December 3; published 2003 December 29

ABSTRACT. We present a method for detecting cosmic rays in single images. The algorithm is based on a simple analysis of the histogram of the image data and does not use any modeling of the picture of the object. It does not require a good signal-to-noise ratio in the image data. Identification of multiple-pixel cosmic-ray hits is realized by running the procedure for detection and replacement iteratively. The tests performed by us show that the method is very effective when applied to the images with spectroscopic data. It is also very fast in comparison with other single-image algorithms found in astronomical data-processing packages. Practical implementation and examples of application are presented.

1. INTRODUCTION

Cosmic-ray hits cause defects in all astronomical images obtained with CCD detectors. A relatively efficient approach to removing traces of cosmic rays from such images is to use multiple frames of the same object and then combine them using an algorithm for rejecting the outlying data. Methods of this type have been presented by, among others, Shaw & Horne (1992) and Windhorst, Franklin, & Neuschaefer (1994). They can be found in most astronomical data-processing packages.

In many cases we are not able to obtain multiple images of the same object, or the required time resolution prevents us from using multiple-image methods. A straightforward method that could be used in such cases would be σ-clipping. In practice it is almost impossible to find a good detection threshold for such an algorithm. In fact, it may leave some obvious cosmic rays untouched while giving numerous spurious detections and rejecting valid data at the same time. Other types of single-image algorithms can also be found in most astronomical data-processing packages. Most of these methods rely on the sharpness of cosmic rays relative to the atmospheric smoothing of real images. These methods are based on some interactive learning techniques (Murtagh 1992; Salzberg et al. 1995) or involve a special fitting of a model representing a real image of a star with a superimposed cosmic ray (Rhoads 2000). These methods were designed to correct direct images of the stellar objects, and we find them to be unsuitable for spectral images. A very efficient method for cosmic-ray detection was presented by van Dokkum (2001). The algorithm based on Laplacian edge detection works well on both imaging and spectroscopic data. It is, however, relatively slow because of the complexity of the approach.

The exposure times of the spectra are usually longer than direct images and range from 600 to 1800 s. A considerable number of cosmic-ray events accumulate in the images during such exposures. We have constructed a simple and straightforward algorithm for detecting cosmic rays in single images. Our method does not need any model of the shape of the image features themselves. We analyze the histograms of pixel counts in small subframes in order to detect pixels deviating by some factor from the bulk of the pixels under consideration.

We would like to note here that low-value bad pixels tend to hide weak cosmic rays by artificially making the local variance larger than would be caused by the usual sources of noise in the image: readout noise, dark current, and Poisson noise associated with background and object signals. The cosmic rays are randomly distributed in the images. On the other hand, bad pixels can be found as deviating pixels at the same position in every frame taken with a given instrument. Therefore, we suggest making bad-pixel correction a separate step of the reduction process prior to cosmic-ray correction.

We describe details of our algorithm in § 2. In § 3 we describe an implementation and the results of our tests. In§ 4 we present examples of application to real astronomical data. A summary is given in § 5.

2. THE ALGORITHM

The proposed algorithm is based on the idea that cosmic rays deposit a portion of their energy in the pixels they hit, causing some extra signal in these pixels. The signal coming from cosmic rays does not have a Gaussian distribution. This should reflect in the distribution of counts in the image affected by cosmic rays. The image as a whole may have a large range

of signal levels in different areas. We analyze relatively small subframes to work with a more concentrated local distribution of counts. In most cases this distribution should be rather compact. The cosmic rays appear as points standing out in the high-count interval.

The proposed algorithm of cosmic-ray detection generally consists of the following steps:

1. Select small sized subframes that cover the whole frame, with substantial overlap.

In each subframe:

2. Calculate the standard deviation of the distribution of counts:

$$\sigma = \sqrt{\frac{\sum c_i^2 - (\sum c_i)^2/n}{n}} \qquad (1)$$

3. Apply a single σ-clipping step to correct the estimate of standard deviation for outlying pixels.

4. Construct a histogram of the distribution of counts.

5. Find the mode of the distribution of counts (i.e., the peak of the histogram).

6. In the interval of counts higher than the mode, find gaps in the histogram (i.e., bins with zero data points).

7. Find the first gap that is wider than a threshold, which is the standard deviation multiplied by an arbitrary number (usually 3.0).

8. If such a gap exists, flag pixels with counts lying above the gap as affected by cosmic rays.

Figure 1 illustrates an example of a histogram with a few pixels identified as affected by cosmic-ray events.

The next step consists of a replacement of the count numbers in the flagged pixels. If we consider a single image, the information about the real signal in the affected pixels is lost. In many cases, however, the characteristic scale of spacial variations of signal are of the order of at least a few pixels. In such cases, one can use some interpolation to replace missing pixel counts. In our implementation we decided to substitute the cosmic rays with the average of the counts in the neighboring pixels.

Cosmic rays are often multiple-pixel events. High-signal pixels may screen neighboring pixels from detection. For this reason we run the procedure of detection and cleaning cosmic rays iteratively. The process is rapidly converging, and usually there are no new detections after two or three iterations. We have also introduced a parameter called the "growing radius," which tells the procedure to clean the pixels within this radius from the detection, even when they do not possess enough signal to be flagged as affected by cosmic rays.

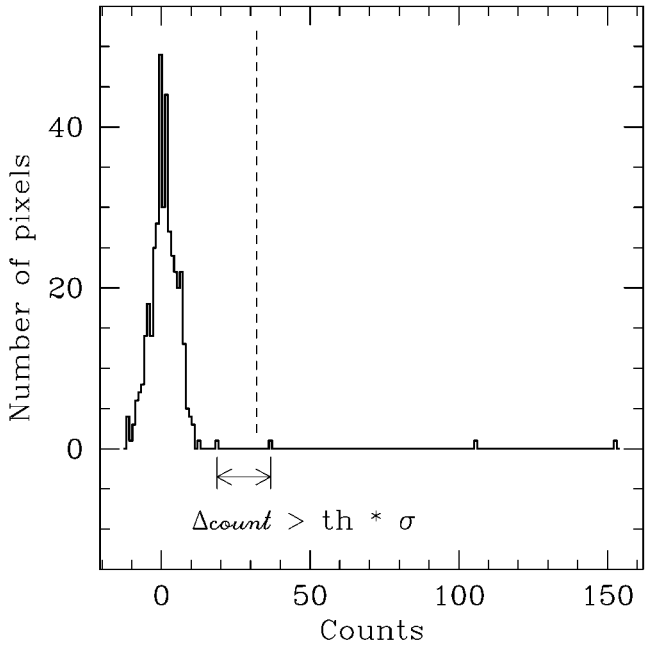

FIG. 1.—Histogram used in a typical application of the cosmic-ray detection algorithm. The vertical dashed line marks the lower limit of counts regarded as the cosmic rays. The arrows mark the first gap in the histogram with a width larger than the threshold.

3. IMPLEMENTATION AND TEST

We have written a computer program that implements the algorithm described above.[1] This program has several input parameters that allow the user to control the process of detecting and cleaning cosmic rays. These parameters define:

1. the size of the subframe box: X and Y dimensions of the rectangle;

2. the threshold: the number by which the local variance is multiplied;

3. the dispersion axis, the interpolation for bad-pixel substitution is calculated along this axis; if no axis selected, it is calculated in the annulus around that pixel;

4. the lower and upper radii of the annulus for the interpolation of data; and

5. the growing radius, the maximum distance of the pixels to be corrected in the neighborhood of the flagged pixel.

The above parameters can be adjusted experimentally to obtain the best results for a given set of data. Our suggestions, however, are as follows. The subframe box should be big enough to contain at least 100 pixels to provide meaningful statistics of counts. On the other hand, it should be small enough to avoid mixing portions of the image that have very different count levels. The low limit of the subframe box size could be

[1] The C source of our program can be downloaded from http://www.camk.edu.pl/~pych/.

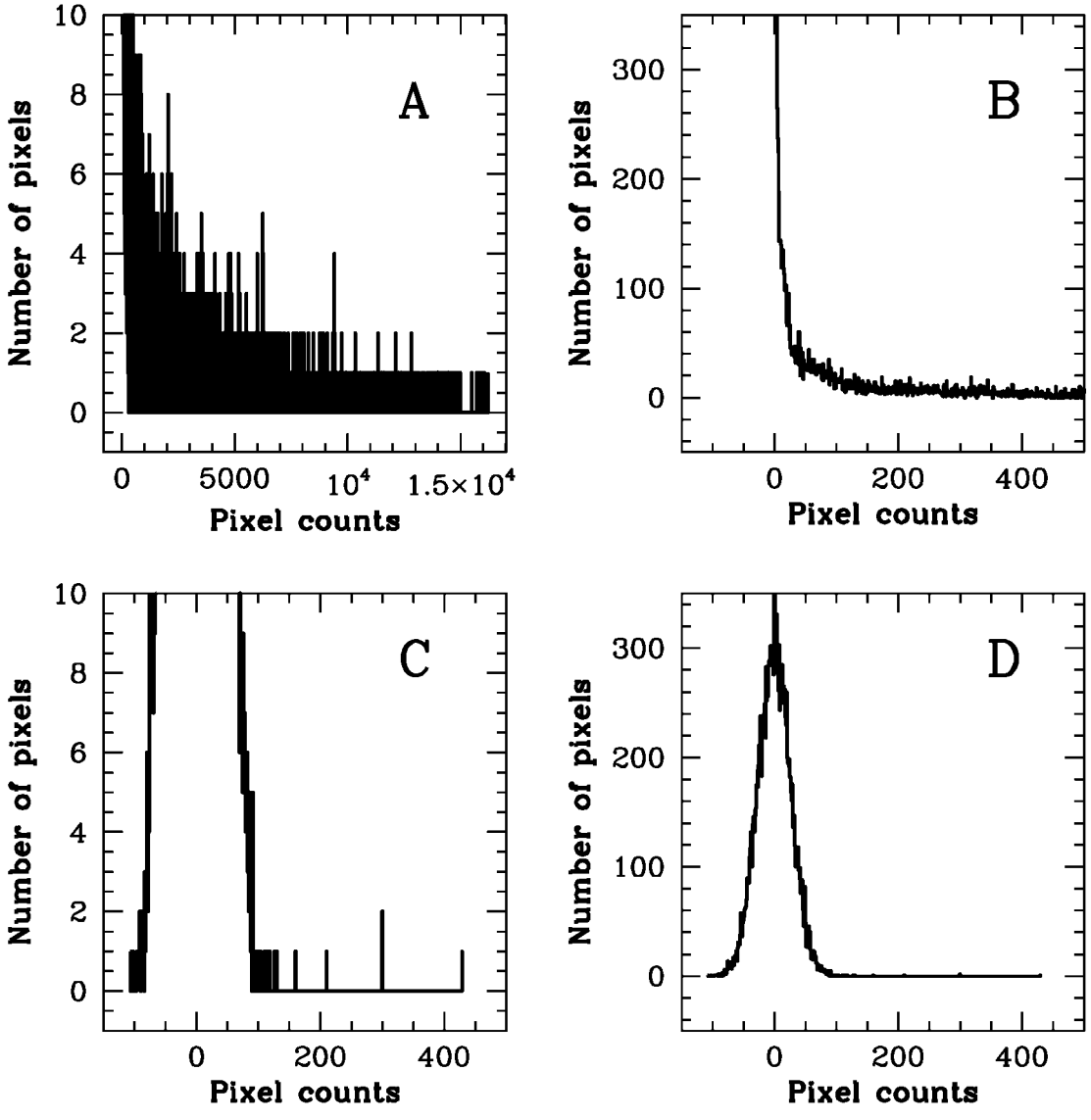

FIG. 2.—Distribution of counts associated with cosmic rays. (*a*) Distribution of counts associated with high-energy cosmic rays. (*b*) Distribution of counts associated with low-energy cosmic rays. (*c*) Distribution of counts associated with undetected cosmic rays. (*d*) Distribution of residual counts after cosmic-ray cleaning. (See text for details.)

the main limitation for the analysis of the imaging frames. The threshold should not be much smaller than 3.0, to avoid erroneous modification of the real data. On the other hand, it should not be much larger than about 3.0, to assure high detection efficiency. The lower and the upper interpolation radii should be set according to the characteristic scale of the variability in the image. Our experience is that in most of the cases it is advisable to set the growing radius parameter to 1. Sometimes neighboring pixels are not affected, and 0 works well. Usually it is not advisable to set a larger number here.

Our program creates a map of detected cosmic rays. This map can be examined to check for any correlation with the real data. The existence of such correlation indicates that the detection threshold has been set too low and the real data has been modified.

To check the capabilities of our method, we have generated an artificial two-dimensional echelle spectrum 2048 × 4096 pixels in size. For this, we have used a procedure from IRAF package *noao.artdata.mkechelle*, with its default parameters. We have also used the procedure *noao.artdata.mknoise* to obtain two noise images. The first one consists of readout noise (5 electrons) and a background of 500 counts with Poisson noise. The second image consists of 1000 cosmic rays with a maximum energy of 30,000 electrons.

Our program finished with zero detections on the noiseless image of the echelle spectrum. The result for the spectrum with

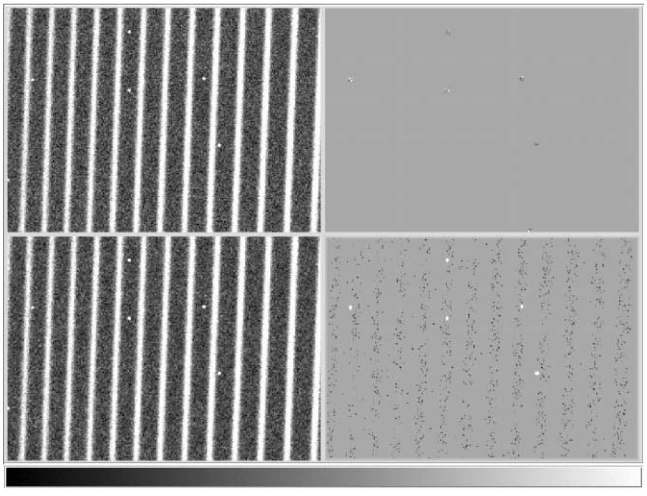

FIG. 3.—Central fragments of our test images. The upper left panel shows the artificial echelle spectrum with the Poisson noise and cosmic rays added. The upper right panel shows the difference between the image cleaned by our program and the original spectrum (without cosmic rays). The lower left panel presents the same image as in the panel above, but after being processed by the IRAF procedure *noao.imred.crutil.cosmicrays*. The lower right panel shows the difference between the image obtained from the IRAF routine *noao.imred.crutil.cosmicrays* and the original spectrum (without cosmic rays).

the readout and Poisson noise added was one detection over the whole frame. Finally, we tested the procedure on a spectrum with Poisson noise and cosmic rays added to the original image of a spectrum. Our program found all but one cosmic ray event of peak above 200 counts. The only event left, with a peak of 430 counts, was located in the vicinity of a trace of the spectrum and was concealed by this signal.

Figure 2 presents histograms of the count distribution. Figure 2*a* presents the original distribution of counts from cosmic rays. The *Y*-axis is cut off at a value of 10 pixels, to highlight the distribution of high-energy hits. Figure 2*b* presents the distribution of cosmic rays in the interval of small counts. Figures 2*c* and 2*d* present the distribution of counts in the image obtained as a difference between the spectrum image after cosmic-ray cleaning and the image without cosmic rays added. The *Y*-axis in Figure 2*c* is cut off at a value of 10 pixels, to show the distribution of pixels with undetected cosmic rays. The features visible above 200 counts are the traces of a single undetected cosmic-ray event. Figure 2*d* presents the distribution of the residual counts. It has the shape of a Gaussian with a small "high end" tail produced by undetected cosmic rays. The standard deviation of the residuals, $\sigma = 26.0$, is similar to the variance of the original image without cosmic rays, $\sigma = 27.8$. This similarity reflects the nature of the adopted substitution method for the pixels affected by cosmic rays, which is simply a mean value of the surrounding pixels.

Our tests were conducted on a PC with a 500 MHz Intel Celeron processor. The program was compiled using a GNU C compiler (gcc ver. 3.2.2) running Red Hat Linux 9.0. The

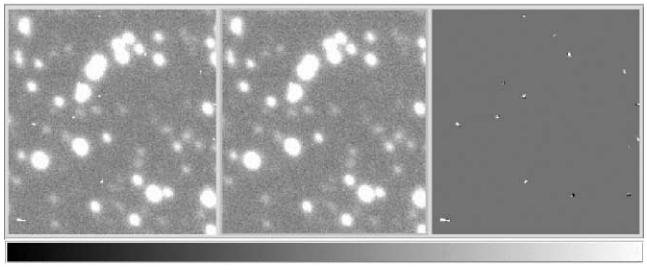

FIG. 4.—Example of an application of the method to a frame of a stellar-field image before (*left*) and after (*middle*) cosmic-ray cleaning. *Right*: map of the cosmic rays.

CPU time needed to process a single frame (2048 × 4096 32 bit pixels) was about 40 s. The CPU time required to process an image depends linearly on the number of image pixels.

We have also run the IRAF procedure *noao.imred.crutil.cosmicrays*. The CPU time needed to process the same image was over 900 s. Compared with the numbers above, our algorithm may be classified as a fast one. The relative speed results from the fact that the whole detection process does not require extensive calculations. Our test demonstrated that the IRAF routine was not able to remove multiple-pixel cosmic-ray events. They were slightly modified, but most of them remained in the frame. At the same time, a large number of image pixels unaffected by cosmic rays were modified. We would like to stress here that our algorithm effectively removes cosmic rays while leaving almost all of the image data untouched.

Figure 3 presents central fragments of our test images. The upper left panel shows the artificial echelle spectrum with the Poisson noise and cosmic rays added. The upper right panel presents the residual signal remaining after subtraction of the original spectrum (without cosmic rays) from the image cleaned by our program. The lower left panel presents the image after being processed by the IRAF procedure *noao.imred.crutil.cosmicrays*. Note that the cosmic rays were not removed. The lower right panel shows the residual signal remaining after subtraction of the original spectrum (without cosmic rays) from the image obtained from the IRAF routine *noao.imred.crutil.cosmicrays*. The scales of the images of the residual signals are the same.

4. EXAMPLES OF PRACTICAL APPLICATION

We have applied the above algorithm to several data sets. Our general experience is that the method works very well on images of both long-slit and echelle spectra. The algorithm was originally designed to work with the spectroscopic data. Our method could be used to clean images of stellar fields with a very wide point-spread function. It works well in a sense that it does not produce too many spurious detections.

The example result of the application to an image of a stellar field in the Large Magellanic Cloud obtained with a CCD cam-

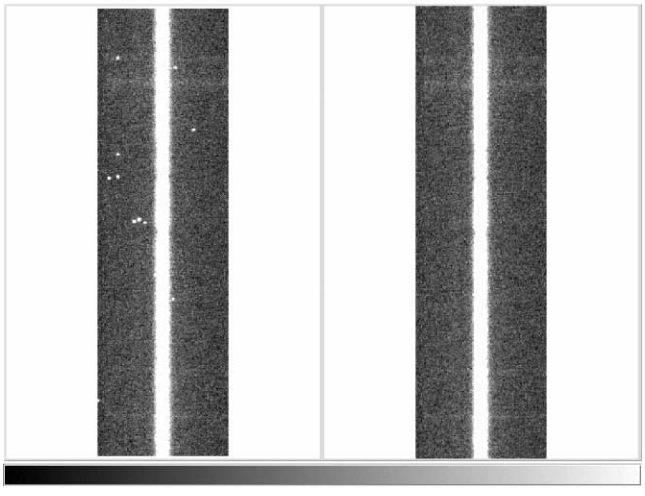

FIG. 5.—Example of a frame with a long-slit spectrum before (*left*) and after (*right*) cosmic-ray cleaning. The stellar signal is in the white band along each panel.

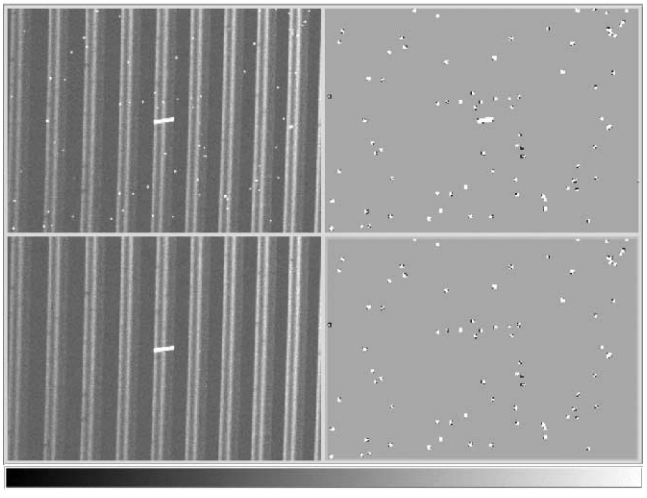

FIG. 6.—Example of the application of our method on an echelle spectrum of a faint star. The upper left panel shows a fragment of the spectrum of a 17 mag star obtained in an 1800 s exposure with the MIKE spectrograph at the Las Campanas Observatory. The upper right panel presents the map of cosmic rays detected in the image. A sky emission line is visible in the center. The lower left panel presents the same fragment of the spectrum as shown in the upper left panel, but after subtraction of the map of cosmic rays modified to remove the sky emission line. The lower right panel shows the map of cosmic rays with the sky emission line removed.

era attached to the 6.5 m Baade Telescope at the Las Campanas Observatory is presented in Figure 4. The atmospheric seeing of about 0″.6 corresponds to the full width at half-maximum (FWHM) of the images of stars that appear slightly larger than 8 pixels.

The limitation of the presented method for these types of images comes from the fact that direct images usually have large count variations within small scales. This produces a large standard deviation of the counts and prevents cosmic rays in the neighborhood of bright objects from being detected. We suggest using one of the methods proposed by Rhoads (2000) or van Dokkum (2001) in such cases.

Figure 5 illustrates a sample result of cosmic-ray removal in a long-slit spectrum. A fragment of the spectrum of a 9.2 mag variable star, obtained in a 1200 s exposure with the 1.88 m telescope at the University of Toronto David Dunlap Observatory, is shown. The left panel presents the original image (after bias-subtraction and flat-field correction) while the right panel shows the cleaned image. No cosmic rays are identified by eye in the cleaned frame.

Our method works best on low-signal images. For the faint stars, however, we encounter another problem. The night-sky emission lines, which are sharp features perpendicular to the dispersion axis, may sometimes be strong enough to be identified as cosmic rays. Our solution to this problem is to edit the map of cosmic rays, erase features identified as sky emission lines (replace the counts with zeros), and then subtract a modified cosmic-ray map from the original image.

Figure 6 illustrates an example of the application on an echelle spectrum of a faint star. The upper left panel shows a fragment of the echelle spectrum of a 17 mag star obtained in an 1800 s exposure with the Magellan Inamori Kyocera Echelle (MIKE) spectrograph at the Las Campanas Observatory. The upper right panel presents the map of cosmic rays detected in the image. A sky emission line is visible in the center. The lower left panel presents the same fragment of the spectrum as shown in the upper left panel, but after subtraction of the modified map of cosmic rays. The map was edited to remove the sky emission line. The lower right panel shows the map of cosmic rays with the sky emission line removed.

5. SUMMARY

We have presented a cosmic-ray rejection algorithm based on a simple analysis of the histogram of the image data. The most important advantage of our method is that it does not require modeling of the image data and may be applied a priori to any type of well-sampled image data. We have checked that for the spectroscopic images it is very effective in detecting cosmic rays, while avoiding numerous spurious detections. Our method does not require advanced and extensive computations, so it is relatively fast.

The weak point of this approach is that bright objects may shield cosmic rays in their neighborhood from detection. This is caused by two factors: first, the sensitivity to cosmic-ray events is reduced at the locations of bright objects because of the Poisson noise associated with the image photons; and second, we look for bright spots, and the legitimate object may be brighter than a nearby comic-ray trace. Our suggestion is to use the algorithm presented above for spectroscopic data whenever multiple-image methods cannot be employed for cosmic-ray removal. A future improvement of the presented

method could be an introduction of better interpolation for the replacement of pixels affected by cosmic rays.

The author would like to thank Prof. S. Rucinski, J. R. Thomson, and H. DeBond for their help in preparing this article. This article was prepared when W. Pych held the NATO Postdoctoral Fellowship administered by the Natural Sciences and Engineering Council of Canada (NSERC). The author also acknowledges the support from the Polish grant KBN 2-P03D-029-23 and the NSERC research grant to S. Rucinski.

REFERENCES

Murtagh, F. D. 1992, in ASP Conf. Ser. 25, Astronomical Data Analysis, Software, and Systems I, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes (San Francisco: ASP), 265
Rhoads, J. E. 2000, PASP, 112, 703
Salzberg, S., Chandar, R., Ford, H., Murthy, S. K., & White, R. 1995, PASP, 107, 279
Shaw, R. A., & Horne, K. 1992, in ASP Conf. Ser. 25, Astronomical Data Analysis, Software, and Systems I, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes (San Francisco: ASP), 311
van Dokkum, P. G. 2001, PASP, 113, 1420
Windhorst, R. A., Franklin, B. E., & Neuschaefer, L. W. 1994, PASP, 106, 798