Application of a multidimensional wavelet denoising algorithm for the detection and characterization of astrophysical sources of gamma rays

S.W. Digel, B. Zhang, J. Chiang, Jalal M. Fadili, Jean-Luc Starck

▶ To cite this version:

HAL Id: hal-00261082
https://hal.archives-ouvertes.fr/hal-00261082

Submitted on 6 Mar 2008

APPLICATION OF A MULTIDIMENSIONAL WAVELET DENOISING ALGORITHM FOR THE DETECTION AND CHARACTERIZATION OF ASTROPHYSICAL SOURCES OF GAMMA RAYS

S. W. DIGEL

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309, USA
E-mail: digel@stanford.edu

B. ZHANG

Quantitative Image Analysis Unit URA CNRS 2582, Institut Pasteur, 25-28, Rue du Docteur Roux, 75724 Paris Cedex 15, France
E-mail: bzhang@pasteur.fr

J. CHIANG

Joint Center for Astrophysics/Physics Department, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD21250, USA & SLAC
E-mail: jchiang@slac.stanford.edu

J. M. FADILI

GREYC CNRS UMR 6072, Image Processing Group, 6, Bd Marechal Juin,14050 Caen Cedex, France
E-mail: Jalal.Fadili@greyc.ensicaen.fr

J.-L. STARCK

Service dÁstrophysique CEA/Saclay, Orme des Merisiers, Bat 709, 91191 Gif-sur-Yvette Cedex, France & Department of Statistics, Stanford University
E-mail: jstarck@cea.fr

Zhang, Fadili, & Starck have recently developed a denoising procedure for Poisson data that offers advantages over other methods of intensity estimation in multiple dimensions. Their procedure, which is nonparametric, is based on thresholding wavelet coefficients. The restoration algorithm applied after thresholding provides good conservation of source flux. We present an investigation of the procedure of Zhang et al. for the detection and characterization of astrophysical sources of high-energy gamma rays, using realistic simulated observations with the Large Area Telescope (LAT). The LAT is to be launched in late 2007 on the Gamma-ray Large Area Space Telescope mission. Source detection in the LAT data is complicated by the low fluxes of point sources relative to the diffuse celestial background, the limited angular resolution, and the tremendous variation of that resolution with energy (from tens of degrees at ∼30 MeV to 0.1° at 10 GeV). The algorithm is very fast relative to traditional likelihood model fitting, and permits immediate estimation of spectral properties. Astrophysical sources of gamma rays, especially active galaxies, are typically quite variable, and our current work may lead to a reliable method to quickly characterize the flaring properties of newly-detected sources.

1. Introduction

The high-energy gamma-ray sky will be studied with unprecedented sensitivity by the Large Area Telescope (LAT) to be launched by NASA on the GLAST mission in late 2007. The catalog of gamma-ray sources from the previous mission in this energy range, EGRET on the Compton Gamma-Ray Observatory, has approximately 270 sources [1]. For the LAT, several thousand gamma-ray sources are expected to be detected, with much more accurately determined locations, spectra, and light curves.

We would like to reliably detect as many celestial sources of gamma rays as possible, and to recognize when a known source is varying or when a faint, previously unknown source flares up to a detectable level. The time scales of flares can be minutes to weeks. In general we will not know in advance where on the sky the sources are that we will detect; projections are that the LAT will detect several times more blazars (a class of active galaxy) than are currently known from observations at other wavelengths.

The fluxes of celestial gamma rays are low, especially relative to the ~ 1 m^2 effective area of the LAT (by far the largest effective collecting area ever in the GeV range). An additional complicating factor is that diffuse, celestial background from the Milky Way itself (which originates in cosmic-ray interactions with interstellar gas and radiation) makes a relatively intense, structured background emission, on which the point sources of interest are superposed. Point sources are so called because they are spatially unresolved. The few very brightest gamma-ray sources will provide approximately 1 detected gamma ray per minute when they are in the field of view of the LAT. The celestial background of the Milky Way will provide about 2 gamma rays per second, distributed over the ~ 2 sr field of view.

For previous high-energy gamma-ray missions, the standard method of source detection has been model fitting — maximizing the likelihood function while moving trial point sources around in the region of the sky being analyzed. This approach has been driven by the limited photon counts and the relatively limited resolution of gamma-ray telescopes. However, at the sensitivity of the LAT even a relatively 'quiet' part of the sky may have 10 or more point sources close enough together to need to be modeled simultaneously when maximizing the (computationally expensive) likelihood function. For this reason, and because we would like to be able to detect spatially resolved sources that do not necessarily have simple shapes (such as a supernova remnant interacting with a gas cloud), non-parametric algorithms for detecting sources are being investigated.

The new wavelet denoising procedure by Zhang, Fadili, & Starck[2] and its application to simulated LAT data are described in the sections that follow.

2. Characteristics of the Data

The LAT (Fig. 1) is a photon-counting detector, converting gamma rays into positron-electron pairs for detection. The trajectories of the pair are tracked and their energies measured in order to reconstruct the direction and energy of the gamma ray.

The energy range of the LAT is very broad, approximately 20 MeV – 300 GeV. At energies below a few hundred MeV, the reconstruction and tracking efficiencies are lower, and the angular resolution is poorer, than at higher energies. The PSF width
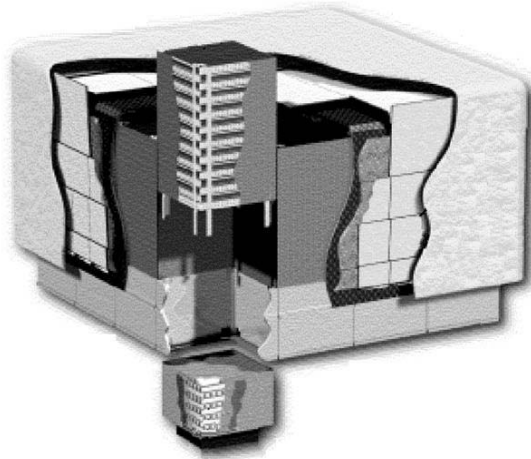

Fig. 1. Cutaway view of the LAT. The LAT is modular; one of the 16 towers is shown with its tracking planes revealed. High-energy gamma rays convert to electron-positron pairs on tungsten foils in the tracking layers. Their trajectories are measured very precisely using silicon strip tracking layers and the energies are determined with the CsI calorimeter at the bottom. The array of plastic scintillators that cover the towers provides an anticoincidence signal for cosmic rays. The outermost layers are a thermal blanket and micrometeoroid shield. The overall dimensions are $1.8 \times 1.8 \times 0.75$ m.

varies from about 3.5° at 100 MeV to better than 0.1° (68% containment) at 10 GeV and above. Owing to large-angle multiple scattering in the tracker, the PSF has broad tails; the 95%/68% containment ratio may be as large as 3.

3. Wavelet Poisson intensity estimation: Overview

Wavelet denoising of LAT data has application as part of an algorithm for quickly detecting celestial sources of gamma rays. The fundamental inputs to high-level analysis of LAT data will be energies, directions, and times of the detected gamma rays. (Pointing history and instrument live times are also inputs for exposure calculations.) For the analysis presented here, we consider the LAT data for some range of time to have been binned into 'cubes' $v(x, y, E)$ of spatial coordinates and energy, because, as we shall see, the wavelet denoising can be applied in multiple dimensions, and so permits estimation of counts spectra. The motivations for filtering data with Poisson noise in the wavelet domain are well known — sources of small angular size are localized in wavelet space.

Many wavelet filtering methods have been devel-

oped based, e.g., on transformations (including variance stabilizing), direct filtering (like Wiener filtering in the wavelet domain), and Bayesian approaches; see Zhang et al.[2] for a review. The recent results of Zhang et al. on wavelet filtering of Poisson data extend the work of Kolaczyk[3] on filtering via hypothesis testing. The appeal of hypothesis testing methods is that they allow quantitative control of significance.

4. Denoising via Wavelet Domain Hypothesis Testing

4.1. *Formulation of the Method*

The background intensity, if not known *a priori*, is assumed to be constant over the region of the sky being analyzed. The background can be estimated from the approximation coefficients at a coarser scale.

For each wavelet coefficient w, the null hypothesis is H_0: w is consistent with the background; and the alternative hypothesis is H_1: w is inconsistent with the background. In the filtering, the H_0 coefficients, which correspond to consistency with the background within statistical fluctuations, are zeroed. The H_1 coefficients, representing regions of wavelet space with significant change from the background, are retained. This is a controlled (via the user-specified p-value) hard thresholding scheme. The coefficients are tested separately. The probability of false detection, i.e., false passing of H_1, is upper bounded by p

$$ E(\frac{N_{H_0}^{reject}}{N_{H_0}}) \leq p, \tag{1} $$

where $N_{H_0}^{reject}$ is the number of coefficients satisfying H_0 but rejected by the corresponding hypothesis test and N_{H_0} is the total number of coefficients satisfying H_0.

For computational tractability, Zhang et al. use the Haar wavelet, because the pdf of the wavelet coefficient conditioning on H_0 of a Poisson process is known in closed form, a non-central Chi-Square distribution. Zhang et al. also show that the same distributions and thresholds apply to the coefficients of biorthogonal Haar (BH) wavelets, which relatively speaking provide good preservation of regularity.

Using the Fisher normal approximation (see Zhang et al.), a threshold can be derived for every Gaussian significance level. This thresholding
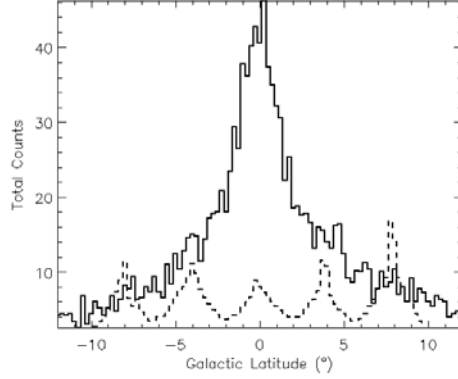

Fig. 3. Solid curve: Latitude profile, averaged over the central 1° of longitude, for the left-hand panel of Fig. 2, i.e., the simulated point sources and celestial background before denoising. Dashed curve: The same profile showing the point-like sources remaining after denoising with a model for the background, corresponding to the right-hand panel of Fig. 2. *This profile has been scaled by a factor of 10 for clarity.*

method has a greater detection power than Kolaczyk's method[3] and Zhang et al.[2] have extended it to non-constant, or unknown *a priori*, backgrounds.

4.2. *Extension to the Energy Dimension*

The hypothesis testing approach can be extended straightforwardly to a third dimension. A 2-dimensional BH transform is applied to each image (one per energy band). To each BH spatial coefficient, a 1-dimensional BH transform is applied along the energy axis, and the hypothesis testing estimator is applied. Then the inverse transformations are applied along the energy axis and spatially to obtain the denoised multi-spectral data.

5. Application to Simulated LAT Data

Figure 2 illustrates the application of the wavelet denoising algorithm for source detection against the bright, structured celestial emission of the Milky Way. This is literally just an illustration with an artificial arrangement of point sources, but the intensity of the celestial background and the flux of the sources are realistic. The region used is in fact more or less a worst case, having the brightest and most strongly varying background for any place on the sky. The center panel of Fig. 2 is the result of requiring the algorithm to estimate the background itself and the right-hand panel shows the result when a background model is provided. The profiles in Fig. 3 show that
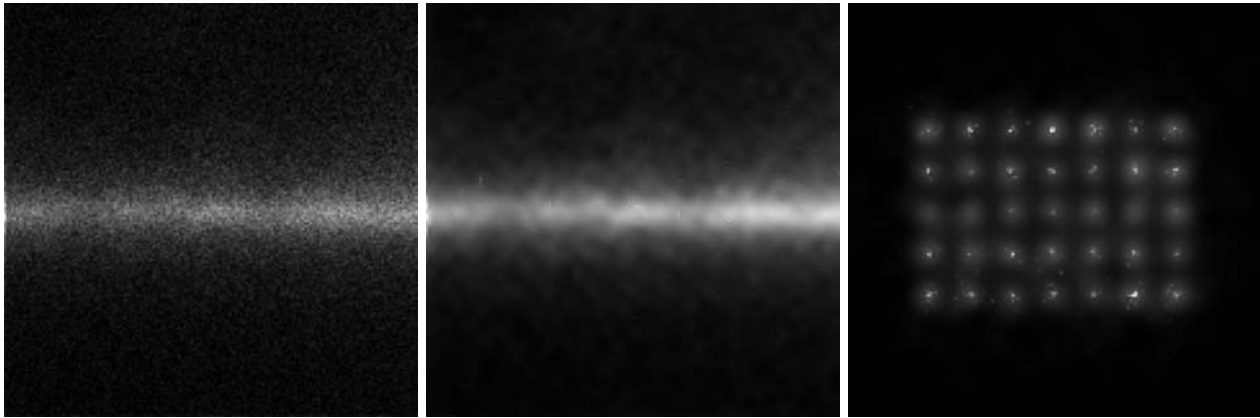
Fig. 2. Example of application of the wavelet denoising algorithm for detecting sources in simulated LAT data. An array of point sources, spaced by 4° was superimposed against the bright celestial background of the inner Milky Way and the exposure from a one-month sky was survey simulated. The point sources are identical, with flux 5×10^{-8} cm^{-2} s^{-1} (>100 MeV) and photon spectral index -2. *(left)* Before denoising; counts summed over energy >100 MeV. *(center)* After application of the denoising algorithm without providing a model for the celestial background. *(right)* After application of the denoising algorithm, which was given a model of the background.
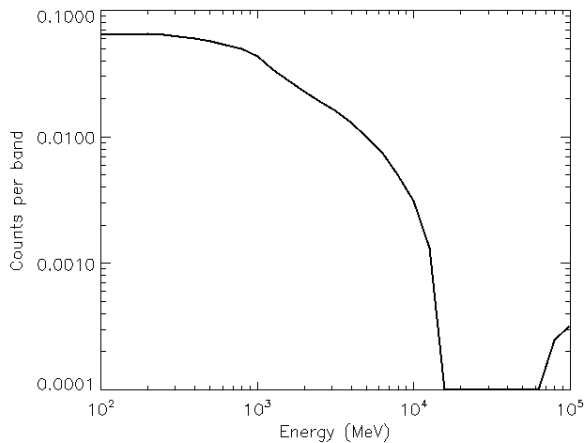

Fig. 4. Example of denoised spectrum from a source in Fig. 2. The distribution of photon counts has approximately the E^{-1} slope expected in counts per band for a differential E^{-2} spectrum.

the bright, structured emission of the Milky Way can totally mask the faint sources; this is why the denoising without a prior background model failed to find the sources. However, the result is quite promising when the denoising is made with a good model for the celestial background gamma-ray emission.

Figure 4 shows a denoised counts spectrum for one of these sources. The spectral characteristics of the denoising have not been studied in detail, but the figure illustrates that the spectral slope is approximately what would be expected for the power-law spectra (photon index -2) of the input sources. Some roll-over is expected owing to the decline of the effective area of the LAT below ∼300 MeV.

6. Summary

We have described the motivations for identifying a reliable nonparametric source detection algorithm to apply to GLAST LAT data. For the relatively short time ranges over which we will want to study sources, the data will be squarely in the low counts regime with widely varying response functions and significant celestial backgrounds. The hypothesis-testing wavelet denoising algorithm of Zhang et al. is a promising algorithm that we are exploring. It can use a model for the background for increased sensitivity. The algorithm also can be applied in the energy dimension and so allows spectra to be recovered.

References

1. R. C. Hartman et al., *Astrophys. J. Supp.* **123**, 79 (1999).
2. B. Zhang, B., J. Fadili, & J-L. Starck, Tech. Rep. CEA-Saclay (2005).
3. E. D. Kolaczyk, *J. Amer. Stat. Assoc.* **94**, 920 (1999).