32ND INTERNATIONAL COSMIC RAY CONFERENCE, BEIJING 2011

A needlet-based approach to the shower-mode data analysis in the ARGO-YBJ experiment

R. IUPPA[1,2], G. DI SCIASCIO[2], F.K. HANSEN[3], D. MARINUCCI[4], R. SANTONICO[1,2]
AND THE ARGO-YBJ COLLABORATION
[1] *Department of Physics, University of Tor Vergata,*
[2]*INFN "Tor Vergata"*
[3] *Institute of Theoretical Astrophysics, University of Oslo*
[4] *Department of Mathematics, University of Tor Vergata*
roberto.iuppa@roma2.infn.it

Abstract: The ARGO-YBJ experiment, located at the Yangbajing Cosmic Ray Laboratory (Tibet, 4300 m asl, 606 g/cm2), is an EAS-array exploiting the full coverage approach at high altitude. The large field of view (2 sr) and the low energy threshold (few hundreds of GeV) result in a trigger rate of ~ 3.5 kHz and $\sim 10^{11}$ EAS collected per year. Such a dataset contains signals laying on different angular scales: point-like and extended gamma-ray sources, as well as large and intermediate scale cosmic-ray anisotropies. The separation of all these contributions is crucial, mostly when they overlap with each other. Needlets are a new form of spherical wavelets that have recently drawn a lot of attention in the cosmological literature, especially in connection with the analysis of CMB data. Needlets enjoy a number of important statistical and numerical properties which suggest that they can be very effective in handling cosmic-ray and gamma-ray data analysis. An unprecedented application to astroparticle physics is shown here. In particular, we focus on their use for background estimation, which is expected to be optimal or nearly-optimal in a well-defined mathematical sense, and for point-source detection. This technique is applied here to the ARGO-YBJ dataset, stressing its advantages with respect to standard methods.

Keywords: Extensive Air Showers, Cosmic Rays, needlets, ARGO-YBJ

Introduction

Over the last decades very high energy astrophysics experiments made great headway in sensitivity and duty-cycle, providing huge amounts of very high quality data. At the present time both space and ground-based telescopes with large field of view and angular resolution as good as few tens of arcminutes are operated. Every day experimenters cope with problems like resolving sources close each other or separating them from diffuse or extended back/foreground or enhancing the significance of the detection, as well as defining the shape of extended sources or giving accurate estimation of directional anisotropies. For instance, ARGO-YBJ observed very intense localized excesses of cosmic rays, $10^\circ - 30^\circ$ wide [1]. The contribution of these regions must be carefully considered when looking at gamma-ray sources nearby. The Cygnus region may present similar problems, because many known gamma-ray sources are there within few degrees.

In the framework of the standard analysis techniques, if the density in a certain direction on the sphere is considered *as it is*, no information can be obtained about the angular scales which the signal comes from in that point. To get *exactly* such information, the harmonic expansion can be used, but in this case no localization in the real space can be obtained anymore. A very good tradeoff is represented by the wavelet expansion, where the exactness of the harmonic expansion is given up in favour of the possibility of having good localization properties in real and harmonic domain at the same time. This is the reason why a growing interest has been devoted in the last five years to the application in a cosmological environment of a new form of spherical wavelets, called *needlets*. Needlets were introduced in the mathematical literature by [2], see also [3] for extensions and generalizations; several applications to Cosmic Microwave Background data data have also been implemented: see for instance [4, 5]. More recently, a few papers have focussed on the use of needlets to develop estimators within the thresholding paradigm, in the framework of directional data, which provide the mathematical formalism for cosmic rays experiments [6].

In this paper, we shall first review briefly the main features of the needlet construction, and explain how its properties make it a very promising tool for cosmic rays data analysis. We then discuss the main features of the ARGO-YBJ experiment, and present applications of needlet procedures both on simulated data and on actual observations. A final section summarizes results and discusses perspectives for further research.

1 Needlets construction and main properties

Let $b(.)$ be a weight function satisfying three conditions, namely $b(t)$ is strictly larger than zero only for $t \in [B^{-1}, B]$, some $B > 1$, $b(t)$ is smooth, and for all $l = 1, 2, ...$ we have $\sum_{j=0}^{\infty} b^2(\frac{l}{B^j}) = 1$. Recipes to construct a function $b(.)$ that satisfy these conditions are provided for instance by [4]. The needlet system is then defined by

$$\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{l=B^{j-1}}^{B^{j+1}} \sum_{m=-l}^{l} b(\frac{l}{B^j}) Y_{lm}(x) \overline{Y}_{lm}(\xi_{jk}) \ ,$$

where λ_{jk} are scaling factors (proportional to pthe pixel area) and ξ_{jk} are pixel centres, as provided for instance by HealPix. The corresponding needlet coefficients are provided by

$$\beta_{jk} = \int_{S^2} f(x) \psi_{jk}(x) dx \ . \tag{1}$$

Needlets enjoy quite a few important properties that make them very suitable for spherical data analysis. Indeed, needlets do not rely on any tangent plane approximation and they are perfectly adapted to standard packages; at each scale j needlets are supported on a finite number of multipoles which are controlled by the data analyst; in real space for any fixed angular distance the tail of the needlets decay faster than any polynomial, i.e. quasi-exponentially as the frequency increases; the following *reconstruction property* holds:

$$f(x) = \sum_{jk} \beta_{jk} \psi_{jk}(x) \ . \tag{2}$$

More recently, the needlet idea has been extended by Geller and Mayeli [3] with the construction of so called Mexican needlets, see also [5] for numerical analysis and implementation in a cosmological framework. Loosely speaking, the idea is to replace $b(\frac{l}{B^j})$ by a smooth function of the form

$$b(\frac{l}{B^j}) = (\frac{l}{B^j})^{2p} exp\left(-\frac{l^2}{B^{2j}}\right) \ , \tag{3}$$

for some integer parameter p. Mexican needlets have localization properties in real space even better than standard needlets, that is why results obtained with prescription (3) are pesented here.

We shall now consider the analysis of data from cosmic rays observatories, following classical approaches to wavelet-based density estimation. Denote by $X_i)$ the observed directions of incoming cosmic rays, and consider the needlet coefficient estimator

$$\widehat{\beta}_{jk} = \sum_{i=1}^{n} \psi_{jk}(X_i) \ . \tag{4}$$

The idea will shall pursue is to implement *thresholding* estimates, as discussed for instance by [6]. More precisely, we can consider the nonlinear estimate

$$\widehat{f}_n^*(x) = \sum_{jk} \widehat{\beta}_{jk} w_{jk}(\widehat{\beta}_{jk}) \psi_{jk}(x) \ , \tag{5}$$

] where $w_{jk}(\widehat{\beta}_{jk})$ is some nonlinear function that ”shrinks” beta towards zero. Such estimates can be shown to be robust and nearly optimal over a wide class of density functions and different loss functions, i.e. figures of merit by which to measure when the estimate is ”close” to the density to be estimated. Hereafter, we will name *source-map* the sky-map as it comes from the experiment; *beta-maps* those containing the coefficient estimators $\widehat{\beta}_{jk}$ (equation4); *reconstructed map* those defined by equation 5. When the needlet transform is numerically computed, three parameters are important, i.e. B (aforementioned), j_0 (the first order computed) and n_j (how many needlet orders are computed). If Mexican needlet are used, also the number p should be specified, though it is usually set to 1.

2 Simulation: suitable set up of the transform

Many figures of merit can be defined to characterize the performance of the needlet transform, mostly looking at the reconstructed signal. Just to give an example, we introduce here two quantities which can be evaluated for simulated signal from point-like sources:

- $\rho_r = N_r^R / N_r^S$, where r is a certain angular distance from the source and R and S stand for “reconstructed” and “sampled” respectively. ρ_r is the ratio of the number of events reconstructed within r from the source to the number of events sampled in the same region;
- $< \delta_r > = \sqrt{\sum_{i=1}^{n_r} \left[(\mathcal{N}_i^R - \mathcal{N}_i^S)/\mathcal{N}_i^S \right]^2 / n_r}$, where the bin index i takes values $[1, n_r]$ corresponding to angular distances from the source $(0, r)$. $\mathcal{N}_i^R$ and $\mathcal{N}_i^S$ indicate the number of events in the i^{th} bin for the “reconstructed” and “sampled” map respectively. $< \delta_r >$ is the square root of the χ^2 of the reconstructed radial distribution to the sampled one, after normalizing it to the bin number. Smaller $< \delta_r >$, higher the confidence that not only the integral, but also the shape of the reconstructed signal reproduces well the sampled one.

We computed ρ_r and $< \delta_r >$ for many combinations of parameters $(B, j_0, n_j$, for both standard and Mexican needlet, as well as for many source extensions and detector point spread functions (PSFs).

Figure 1 reports schematic views of ρ_r and $< \delta_r >$ as functions of j_0 and n_j. To realize it, we simulated a source emitting 10^6 events in the ARGO-YBJ FOV, assuming a gaussian PSF with $\sigma = 1°$ and applied the Mexican needlet transform ($B = 1.6$, $p = 1$, $j_0 = 1$, $n_j = 15$). The radius chosen to evaluate ρ and $< \delta >$ was $r = 3\sigma$ and no background contribution is taken in consideration. For both plots, the first order used to reconstruct the signal is represented on the horizontal axis, while the vertical axis number of orders used. The color scale represents the values of $\rho_{3\sigma}$ and $< \delta_{3\sigma} >$. For instance, bins $(5, 6)$ contain the values of $\rho_{3\sigma}$ and $< \delta_{3\sigma} >$ obtained from the map reconstructed with needlet orders $5 \rightarrow 10$. What is easy noticing

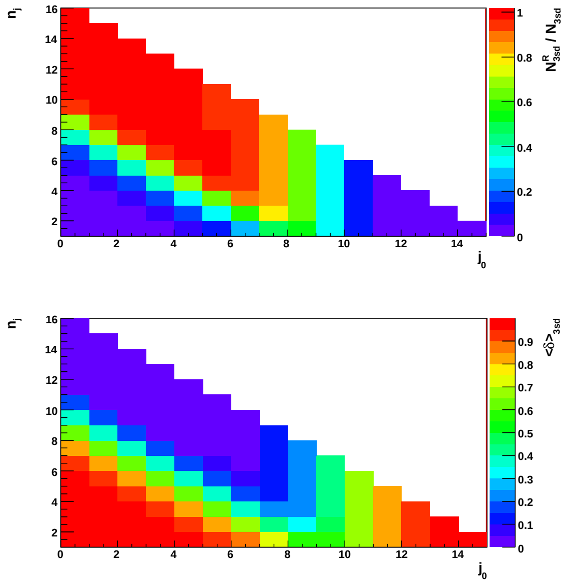

Figure 1: Representation of the reconstruction properties of the Mexican needlet transform. *Upper plot:* $\rho_{3\sigma}$. *Lower plot:* $< \delta_{3\sigma} >$. See text for definition and details.

is that the two plot are somehow conjugate. Where the ratio $\rho_{3\sigma}$ is close to 1, the error $< \delta_{3\sigma} >$ is close to 0 and viceversa. It happens in the top-left part of the schemes, representing cases where almost all the beta-maps computed in the transform are used in the reconstruction. Nonetheless a very good reconstruction can be obtained also with less expensive choices: e.g. if orders $5 - 11$ are used, more than 95% of events will be part of the reconstruction and the signal shape will be good too, because the error is less than 5%.

3 Preliminary analysis of ARGO-YBJ data

The ARGO-YBJ experiment is a wide field of view air shower array located at the YangBaJing Cosmic Ray Laboratory (Tibet, P.R. China, 4300 m a.s.l., 606 g/cm^2). Details about the experimental setup and the event reconstruction can be found in [7] and [8] respectively.

We present here the application of the needlet analysis to data taken by the ARGO-YBJ telescope from November 2007 to September 2010. Since the development of the technique is still at a preliminary stage, no attention has been paid in optimizing data-selection for point-sources or cosmic-rays extended excesses. We asked for as much statistics as possible and selected events firing more than 40 strips on the central carpet. The reconstructed arrival direction is asked to have zenith angle less than $50°$. Selected events are $1.27\,10^{11}$. The background of cosmic rays at angular scales up to $45°$ has been estimated with the Direct Integration method [9]. The source-map, i.e. the difference between the events and the background map, has been built following the HEALPix pixelisation scheme [10]. The needlet transform has been applied to the background-subtracted map, using the NEEDPACK, a fast software tool developed on purpose. Results reported here are about Mexican needlet transform with $B = 1.6, p = 1, j_0 = 1, n_j = 11$.

3.1 β-maps and significance estimation

Figure 2 represents the beta-map obtained for order 5. As we expect from the corresponding angular scale ($(1.6)^5 \sim 17°$), intermediate scale anisotropies [1] are well visible. To evaluate the significance of the signal with respect to

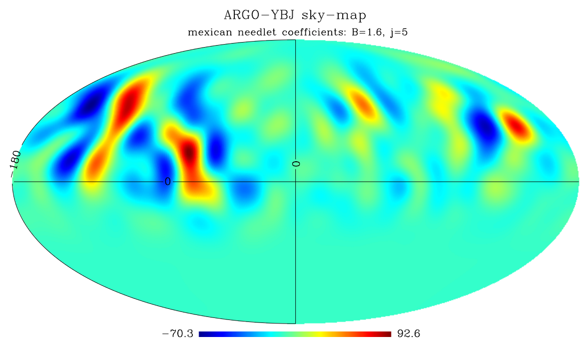

Figure 2: Beta-map from the sky observed by ARGO-YBJ. Mexican needlet transformation with $B = 1.6$, $p = 1$, $j = 5$.

the estimated background, we used the background estimation $\mathcal{B}_k$ as mean map and sampled 1000 random background copies $\mathcal{B}^i_k$, $\mathrm{i} = 1, \ldots, 1000$; every map has been needlet-transformed so that the variance $\sigma^2_{jk} = var(\beta_{jk})$ has been estimated for each pixel k at every order j. These "variance" maps may be directly used to calculate the significance of the content of the k^{th} bin in the j^{th} beta-map: β_{jk}/σ_{jk}. Figure 3 shows the distribution of β_{9k}/σ_{9k}. The curve represents the gaussian fit to the $\hat{\beta}_{9k}/\sigma_{9k}$ distribution, where $\hat{\beta}_{9k}$ is the transform of a background random copy (background subtracted). The fit results are $\mu = -0.0005 \pm 0.0011$ and $\sigma = 1.0014 \pm 0.0021$, as it should be ($\chi^2/\mathrm{d.o.f.} = 38.9/37$). The distribution of the signal map significantly deviates from the gaussian trend. The contribution of the sources is well visible from 3.5 up to 10 s.d. There is also a small excess below -3 s.d., whose origin is due to the shape of the needlet function.

This method allows to estimate the significance of every excess *directly* in the β space, with no need of signal-smoothing or averaging. Moreover, it must be recalled that

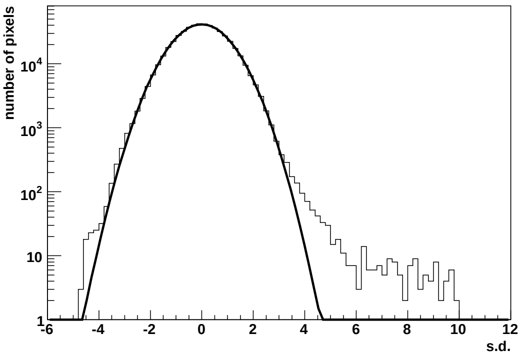

Figure 3: Significance distribution of the pixels of the 9^{th} order beta-map from the sky observed by ARGO-YBJ.

a given signal may be there in two or more j orders and that the significance estimated here refers to the *single-order*.

3.2 Thresholding

The estimation of the variance-maps σ_{jk} may be also used to manipulate the beta-maps before using them to reconstruct the signal. We present here a "soft thresholding" technique, aimed to get rid of the non-significant coefficients. To do that, we introduce the weight functions

$$w_{jk}(\beta_{jk}) = \frac{1}{2}\left(1 + \tanh\frac{\beta_{jk}/\sigma_{jk} - T}{L}\right)$$

, where T is a certain threshold, passing which the weight function goes from 0 to 1. The L parameter is related to the width of the transition region. To give an example of the application of the "soft thresholding" method we set $T = 3$ and $L = 0.2$ for orders $j = 5 - 11$[1] of the Mexican needlet transform ($B = 1.6$ and $p = 1$), then used formula 5 to get the reconstructed map. A zoom of the result around the Crab nebula is given in figure 4. The source is well visible and no fluctuations typical of images from EAS arrays are there. The number of events reconstructed within $5°$ from the nominal source position is $(113 \pm 3)10^3$, to be compared with that obtained from standard analysis techniques (e.g. [11]) $(117 \pm 2)10^3$.

4 Conclusions and perspectives

Recently, needlets drew the attention of the scientific community for their important applications in data-analysis of cosmological data as a new form of spherical wavelets. We presented here the very first application of the needlet transform to high energy astrophysics, showing the very good properties of localization.The needlet transform is sensitive in the whole harmonic domain, provided that enough orders of needlets are computed. In particular, the application to the ARGO-YBJ data-set found again the well-known intermediate scale anisotropy at low orders and that of point gamma-ray sources at high orders, thus showing the possiblity of decting sources directly in the needlet space. The significance of the beta-maps is carried out quite easily if the background distribution is known and the variance of the beta coefficients may be used to threshold the signal in the beta space, which turned out to be a very promising technique.

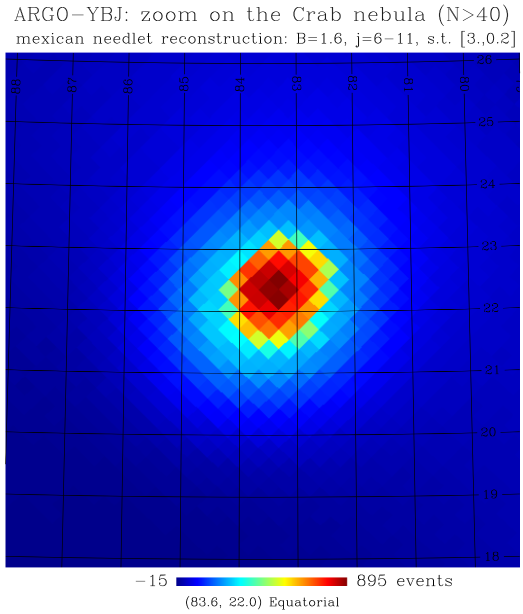

Figure 4: Signal of the Crab region reconstructed with the Mexican needlet technique. Parameters of the transform: $B = 1.6, p = 1, j = 5 - 11$, soft thresholding $[3.0, 0.2]$

References

[1] R. Iuppa et al., these proceedings (507), 2011
[2] F.J. Narcowich et al., SIMA, 2006, **38**: page 574-594
[3] D. Geller et al., Math. Z., 2009, **262**: pages 895-927, arXiv: math/0602201
[4] D. Marinucci et al., MNRAS, 2008, **383**(2): page 539-545, arXiv: 0707.0844
[5] S. Scodeller et al., 2010, preprint, arXiv: 1004.5576
[6] P. Baldi et al., Ann. of Stat., 2009, **37**(6A): page 3362-3395, arXiv:0807.5059
[7] G. Aielli et al., NIMA, 2006, **562**: page 92-96
[8] R. Iuppa et al., Proceedings of the 31^{st} ICRC (432), arXiv:1001.0533
[9] R. Fleysher et al., Apj, 2004, **603**: page 355-362
[10] K.M. Gòrski et al., Apj, **622**(2): page 759-771, arXiv:astro-ph/0409513
[11] B. Bartoli et al., Long-term monitor of Mrk 421 TeV emission using ARGO-YBJ experiment, submitted to Apj

1. We choose these orders because of the PSF of ARGO-YBJ for point-like gamma-ray sources at $N > 40$.