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

G. Di Giacis\textsuperscript{1}, F.K. Hansen\textsuperscript{2}, R. Iuppa\textsuperscript{1,3}, D. Marinucci\textsuperscript{4} and R. Santonico\textsuperscript{1,3}

ON BEHALF OF THE ARGO-YBJ COLLABORATION

\textsuperscript{1}INFN “Tor Vergata”

\textsuperscript{2}Institute of Theoretical Astrophysics, University of Oslo

\textsuperscript{3}Department of Physics, University of Tor Vergata.

\textsuperscript{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/cm\textsuperscript{2}), 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 3.6 kHz and $10^{11}$ 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 whole 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. Recently, wavelet techniques have become a well-established tool for the analysis of cosmological and astrophysical data, see for instance \cite{18, 19} and the references therein. In particular, a growing interest has been devoted in the last 5 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 \cite{12}, see also \cite{6}, \cite{7} for extensions and generalizations. Stochastic properties of needlets when implemented on spherical random fields were first given by \cite{1}, \cite{2}, where applications to several statistical procedures are also considered. Several applications to Cosmic Microwave Background data data have already been implemented: see for instance \cite{14}, \cite{11}, \cite{15}, \cite{16}, \cite{4}, \cite{17}.

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. Thresholding estimates were introduced in the statistical literature by Donoho et al. in \cite{5}; we refer to \cite{8} for a textbook reference. In the mathematical literature, needlet-based thresholding algorithms on the sphere are discussed by \cite{3}, \cite{10}, \cite{9}.

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 technical 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

For functions on the sphere, it is well known that the following spectral representation holds:

$$f(x) = \sum_{lm} a_{lm} Y_{lm}(x), \quad a_{lm} = \int_{S^2} f(x) \overline{Y}_{lm}(x) dx,$$
where the set \{Y_{lm}\} represents the array of so-called spherical harmonics.

In the presence of a partially observed sky, the evaluation of these inverse Fourier transforms becomes unfeasible. Moreover, localization in both the real and the harmonic space is indeed necessary when searching for localized features, such as for instance point sources. In view of these considerations, it is natural to look for system of functions with double localization properties (in real and harmonic spaces), i.e. spherical wavelets. Among spherical wavelets, we shall be concerned with needlets, whose construction we review as follows.

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 \( \sum_{j=0}^{\infty} b^2(\frac{j}{2^l}) = 1 \). Recipes to construct a function \( b(.) \) that satisfy these conditions are provided for instance by \[11\]. The needlet system is then defined by

\[
\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{l=|B^{-1}|}^{B^{-1}+1} \sum_{m=-l}^{l} b\left(\frac{l}{B}\right) Y_{lm}(x) \nabla_{lm}(\xi_{jk}),
\]

where \( \lambda_{jk} \) are scaling factors (proportional to the pixel area) and \( \xi_{jk} \) are pixel centres, as provided for instance by HealPix. The corresponding needlet coefficients are provided by

\[
\beta_{jk} = \int_{S^2} f(x) \psi_{jk}(x) dx.
\]

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:

\[
f(x) = \sum_{jk} \beta_{jk} \psi_{jk}(x).
\]

More recently, the needlet idea has been extended by Geller and Mayeli \([6, 7]\) with the construction of so-called Mexican needlets, see also \[17\] for numerical analysis and implementation in a cosmological framework. Loosely speaking, the idea is to replace \( b(\frac{t}{B}) \) by a smooth function of the form

\[
b\left(\frac{l}{B}\right) = \left(\frac{l}{B}\right)^{2p} \exp\left(-\frac{l^2}{B^2}\right),
\]

for some integer parameter \( p \). Mexican needlets have extremely good localization properties in real space.

We shall now consider the analysis of data from cosmic rays observatories. The idea which we shall discuss here follows from classical approaches to wavelet-based density estimation, as discussed on the real line by \[5, 8\] and many following references. Denote by \( X_i \) the observed directions of incoming cosmic rays, and consider the needlet coefficient estimator

\[
\hat{\beta}_{jk} = \sum_{i=1}^{n} \psi_{jk}(X_i).
\]

The idea will shall pursue is to implement thresholding estimates, as discussed for instance by \[3, 9, 10\]. More precisely, we can consider the nonlinear estimate

\[
\hat{f}_n(x) = \sum_{jk} \beta_{jk} \omega_{jk}(\hat{\beta}_{jk}) \psi_{jk}(x),
\]

where \( \omega_{jk}(\hat{\beta}_{jk}) \) 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 \( \beta_{jk} \) (equation 3); reconstructed map those defined by equation 4. 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:

- \( \rho_r = \frac{N_r^R}{N_r^S} \), where \( r \) is a certain angular distance from the source and \( R \) and \( S \) stand for "reconstructed" and "sampled" respectively. \( \rho_r \) is the ratio of the number of events reconstructed within \( r \) from the source to the number of events sampled in the same region;

- \( \delta_r = \sqrt{\sum_{i=1}^{n} \left[ \frac{(N_i^R - N_i^S)^2}{N_i^S} \right] / \rho_r} \), where the bin index \( i \) takes values \([1, n_r]\) corresponding to angular distances from the source \((0, r)\). \( N_i^R \) and \( N_i^S \) 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 \( \chi^2 \) 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 \( \rho_r \) 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 $\rho_r$ 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 $\sigma = 1^\circ$ and applied the mexican needlet transform ($B = 1.6, p = 1, j_0 = 1, n_j = 15$). The radius chosen to evaluate $\rho$ 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 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 $\rightarrow$ 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 The ARGO-YBJ experiment

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$^2$). It exploits the full coverage with a central carpet $\sim 74 \times 78$m$^2$ made of a single layer of Resistive Plate Chambers with $\sim 93\%$ active area, enclosed by a guard ring partially instrumented to better the angular resolution. It is operated with a duty-cycle higher than 85% since November 2007 with trigger rate intrinsically stable at level 0.5%. The high altitude, as well as the full coverage approach, lower the energy threshold of this EAS array down to few hundreds GeV. The event reconstruction [?] guarantees angular resolution well below the angular scales dealt with in this paper.

4 Preliminary analysis of ARGO-YBJ data

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 $1.27 \times 10^{11}$. The background of cosmic rays at angular scales larger than 45$^\circ$ has been estimated with the Direct Integration method [?]. The source-map, i.e. the difference between the events and the background map, has been built following the Healpix conventions [?]. 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 $B = 1.6, p = 1, j_0 = 1, n_j = 11$.

4.1 Beta-maps at different angular scales

Upper and lower plot of figure 2 represent the beta-maps obtained for order 5 and 9 respectively. Corresponding angular scales are $(1.6)^5 \sim 17^\circ$ and $(1.6)^9 \sim 2.6^\circ$. In the 5th order map intermediate scale anisotropies are well visible [21]. The $g^{th}$ order map shows point-sources instead: Crab nebula and BL Lac Mrk421 are well visible in the left part of the plot. The color scale of these maps has no immediate physical meaning, though the proportionality to the signal intensity is quite evident.

4.2 Significance estimation

To evaluate the significance of the signal with respect to the estimated background, we used the background estimation $B_k$ as mean map and sampled 1000 random background copies $B_k^i, i = 1, \ldots, 1000$; every map has been needlet-transformed so that the variance $\sigma_{jk}^2 = var(\beta_{jk})$ 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: $\beta_{jk}/\sigma_{jk}$. Figure 3 shows the distribution of $\beta_{jk}/\sigma_{jk}$. The curve represents the gaussian fit to the $\beta_{jk}/\sigma_{jk}$ distri-
4.3 Thresholding

The estimation of the variance-maps $\sigma_{jk}$ introduced in section 4.2 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 $j = 5 - 11$ of the mexican needlet transform ($B = 1.6$ and $p = 1$), then used formula 4 to get the reconstructed map. A zoom of the result around the Crab nebula is given in figure 4. The source is well visible

...
5 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 possibility of detecting 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.

We had no room to deal with them in detail, but we cannot help mentioning further applications of the needlet technique to the ARGO-YBJ data analysis, like the blind search in the beta space, which turned out to be a very promising technique.

References

[21] R. Iuppa et al., these proceedings (507), 2011