ICRC 2001

Selection of TeV γ-rays using the Kernel multivariate technique

S. Dunlea[1], P. Moriarty[2], and D. J. Fegan[1]

[1]Department of Experimental Physics, University College Dublin, Belfield, Dublin 4, Ireland.
[2]School of Science, Galway-Mayo Institute of Technology, Dublin Road, Galway, Ireland.

Abstract. The Kernel multivariate analysis technique is optimised to select γ-ray events from ON/OFF observations of the Crab Nebula recorded by the Whipple 10 m Imaging Atmospheric Cherenkov Telescope in January and February 2000. Results are compared with the conventional Supercuts analysis and with a Neural Network analysis. The technique is also applied to ON/OFF data taken on Markarian 421 during spring of 2000. A method to estimate the energy of γ-ray primaries is examined, and a TeV spectrum of the Crab Nebula extracted on this basis.

1 Introduction

The standard technique employed by the Whipple collaboration to discriminate γ-rays from background events uses a number of selection cuts on the standard parameters that characterise an atmospheric Cherenkov image. This technique, known as Supercuts (Punch et al., 1991), has been developed and improved over the past decade. Supercuts implements the simplest method of delineating a parameter space. In this work we examine a more complex selection algorithm, the Kernel technique, and compare its performance to that of Supercuts. We also compare the results to those obtained by a Neural Network analysis (Dunlea (2001) details this approach).

Conventional analysis techniques have classified events, each described by a set of n parameters, by choosing a surface in parameter space. Events on one side of this surface are classified as γ-rays, and events on the other side are classified as background. Supercuts uses a simple multidimensional box constructed by placing fixed boundary limits on each parameter. An alternative approach is to choose a surface formed by the set of all points in parameter space that share a particular value of the likelihood function R, or its

logarithm:

$$\log\left(R\left(p\right)\right) = \log\left(\frac{f_\gamma\left(p\right)}{f_b\left(p\right)}\right), \tag{1}$$

where p represents a point in parameter space, $f_\gamma\left(p\right)$ is the γ-ray probability distribution as a function of position in parameter space, and $f_b\left(p\right)$ is the background probability distribution. An event on one side of this surface (having a likelihood greater than or equal to this critical value of the likelihood function) is classified as a γ-ray, and an event on the other side (with likelihood less than the critical value) is classified as a background event. The calculation of these probability distributions is the core of the Kernel technique. The use of such a technique in TeV γ-ray astronomy has been described previously by Samuelson (1999) and Moriarty and Samuelson (2000). It has also been applied to high energy physics for the detection of the top quark (Holmström and Sain, 1997).

2 Kernel Analysis

Estimates of these probability distributions are derived from a dataset of over 30000 sample γ-ray simulations and an equal number of real background events. The γ-ray Monte Carlo simulations used in this work have energies ranging from 0.2 to 8 TeV distributed with a differential spectral index of -2.4. They are produced by the KASCADE system as implemented at Iowa State University by Mohanty et al. (1998), a derivative of the system described by Kertzman and Sembroski (1994). These simulations are tailored to resemble the real data acquisition process as closely as possible. The same trigger conditions are applied to the simulated data, and noise comparable to that in real data is added. The resulting γ-ray images should be accurate depictions of real γ-ray events.

No simple function can describe the distribution of the γ-ray images nor the distribution of background events. The Kernel technique attempts to estimate these distributions by

Correspondence to: S. Dunlea (shane@ferdia.ucd.ie)

convolving *each* sample point with a point spread function to obtain a smooth continuous approximation to the probability distribution. This is analogous to estimating the electric potential at some point in space due to charges at the sample points, each of which has a potential function like that of the point spread function. The basic kernel estimator may be written as

$$f_\gamma = \frac{1}{N_\gamma h_\gamma} \sum_{i=1}^{N_\gamma} K\left(\frac{p - \gamma_i}{h_\gamma}\right), \tag{2}$$

where p is again a point in parameter space, h_γ is a scale factor, and $\gamma_1, \ldots, \gamma_{N_\gamma}$ are vectors of parameters of the N_γ sample γ-rays. The point spread, or kernel, function, K, can be any scalar function in n-dimensional space (Hand, 1982; Scott, 1992). In this work, a multivariate Gaussian is used as the kernel function,

$$K = \frac{1}{\sqrt{(2\pi)^n |\xi_\gamma|}} e^{-\frac{1}{2}((p-\gamma)/h_\gamma)^\mathsf{T} \xi_\gamma^{-1} ((p-\gamma)/h_\gamma)}, \tag{3}$$

where n is the number of parameters and ξ_γ is the covariance matrix of the γ-ray dataset. The background distribution can be similarly defined. Scott (1992) has shown that if the kernel is a product of univariate Gaussians (one for each dimension), then the scale factor that minimises the mean integrated squared error between an actual distribution and its kernel estimator is given by

$$h_\gamma = \left(\frac{4}{N_\gamma (n+2)}\right)^{1/(n+4)}. \tag{4}$$

The γ-ray probability distribution may now be written as

$$f_\gamma = \frac{1}{N_\gamma h_\gamma \sqrt{(2\pi)^n |\xi_\gamma|}} \sum_{i=1}^{N_\gamma} e^{-\frac{1}{2h_\gamma^2}(p-\gamma_i)^\mathsf{T} \xi_\gamma^{-1} (p-\gamma_i)}. \tag{5}$$

Similarly, f_b can be defined for the background distribution, so the log-likelihood function, $\log(R)$, can now be calculated using Equation 1.

2.1 Reduction of Computational Overhead

Kernel analysis is computationally intensive as *every* event is compared with *every* γ-ray simulation and with *every* background event. The probability distributions f_γ and f_b defined by Equation 5, represent the convolution of the γ-ray simulations and background samples with a point spread (kernel) function. Therefore, the value of the log-likelihood function, $\log(R)$, can be calculated for a lattice of points in n-dimensional parameter space. Values between the nodes of the lattice can be estimated using linear piecewise interpolation. This process results in a factor of 1400 decrease in the time required to analyse a typical data file. However, while producing the required lattice requires many more calculations than a typical full Kernel analysis, it need only be carried out once per detector configuration.

3 Intercomparison of Neural Network, Kernel, and Supercuts selection strategies

In order to optimise the analyses described in previous sections, a database of 31 ON/OFF pairs taken on the Crab Nebula was established. Runs were taken during January and February 2000, on nights with particularly clear skies. The traditional method of testing a new technique (by optimising the significance of the detection of a signal above background) is followed here. The parameter set (*length*, *width*, *distance*, *alpha*, $\log(size)$), used in this work for both the Neural Network and Kernel techniques, has proven to optimise the capacity to discriminate between γ-rays and background events (see Moriarty et al. (1997) for definitions of these parameters).

Table 1 shows the optimised significances obtained for the Neural Network and Kernel techniques, along with the percentage of the γ-ray simulations that pass each optimised cut. The significance and percentage of passing γ-rays for Supercuts 2000 is shown also for comparison.

Technique	Significance	Rate	% sims
Neural Network	22.32σ	$2.29\ \gamma\ \mathrm{min}^{-1}$	29.6%
Kernel	30.13σ	$2.45\ \gamma\ \mathrm{min}^{-1}$	27.9%
Supercuts 2000	22.18σ	$2.63\ \gamma\ \mathrm{min}^{-1}$	28.8%

Table 1. Results of Crab Nebula data at self-optimised cuts

Detection significance of the Neural Network technique performs similarly to Supercuts, while the Kernel technique performs considerably better. This improvement does not appear to be at the expense of the rate, suggesting an improvement in the rejection of background events.

Table 2 presents the results of 25 ON/OFF pairs of independent Markarian 421 data taken between January and May 2000 using the cuts optimised on the Crab Nebula data.

Technique	Significance	Rate
Neural Network	13.71σ	$1.48\ \gamma\ \mathrm{min}^{-1}$
Kernel	18.67σ	$1.56\ \gamma\ \mathrm{min}^{-1}$
Supercuts	15.73σ	$1.98\ \gamma\ \mathrm{min}^{-1}$

Table 2. Results of Markarian 421 data at optimised cuts

The relative performance of each technique is not particularly different than when applied to the Crab Nebula. The Neural Network and Supercuts techniques result in similar significances but the Neural Network cannot reproduce the same rate. The Kernel technique again realises the highest significance in the ON/OFF mode, while still preserving a reasonably high rate. In contrast to the Crab Nebula results, the Markarian 421 results are independent of the optimisation process, and so provide an unbiased comparison of the different techniques.

In an effort to understand the distribution of events selected by each technique, 2-dimensional scatter plots showing events that pass cuts were drawn. This analysis was per-

formed upon 5000 γ-ray simulations to ensure that a large population of events would pass the cuts. As an example Fig. 2 shows the *length* vs. $\log(size)$ distribution of all these simulations followed by the distributions of events *passing* the respective techniques. The complete 5-dimensional distribution cannot be represented on paper, and while the 2-dimensional plots are adequate for the present discussion, it should be remembered that points which coincide on these plots may be widely separated in 5-dimensional space.

It is clear that two distinct populations exist in the distribution of $\log(size)$ for the Neural Network and Kernel techniques. The larger events coincide with that region selected by Supercuts, but a curious region of smaller events also exists. This suggests that γ-ray events with an intermediate *size* are difficult to distinguish from background events. Thus, few events in this region will pass the optimised cut. The close resemblance of the selection regions of the Neural Network and Kernel techniques endorses the merit of such a selection. The sharp boundaries seen in the Supercuts selection region are evidence of the empirical cut on each parameter, rather than the more elegant single cut of the other techniques.

4 Energy estimation and Spectral analysis

To train a Neural Network to estimate the energy of selected γ-ray events, the standard parameter set of the γ-ray simulations was used as the input to the network. The target output was chosen to be the base-ten logarithm of the energy.

To obtain the γ-ray probability distribution using the Kernel analysis, each event must be compared with every γ-ray simulation. In effect Equation 5 quantifies how similar each γ-ray simulation is to the event and calculates the average. This process naturally lends itself to the estimation of the energy of selected events. The energy of a simulated γ-ray can be used to weight the calculation of its contribution to the probability distribution. Thus, the new weighted probability distribution should be the convolution of the estimated energy of the event with the probability distribution. The estimated energy may then be extracted by dividing this result with the original probability distribution,

$$\tilde{E} = \frac{\sum_{i=1}^{N_\gamma} E_i \, \exp\left(-\frac{1}{2h_\gamma^2} \left(p - \gamma_i\right)^\mathsf{T} \xi_\gamma^{-1} \left(p - \gamma_i\right)\right)}{\sum_{i=1}^{N_\gamma} \exp\left(-\frac{1}{2h_\gamma^2} \left(p - \gamma_i\right)^\mathsf{T} \xi_\gamma^{-1} \left(p - \gamma_i\right)\right)}, \quad (6)$$

where E_i is the energy of the simulated γ-ray event γ_i. Applying this formula to the γ-ray simulations themselves gives good energy reconstruction if the energy estimate, $\tilde{E}$, is modified to $\tilde{E}' = 1.02\tilde{E}^{1.12}$.

Once an energy has been estimated for each candidate γ-ray, the energy spectrum can be derived using the method prescribed by Mohanty et al. (1998). For spectral analyses it is desirable to have as large a population of candidate γ-rays as possible, while still maintaining a very significant detection. Thus, a cut which is optimised purely on the basis of significance is often too restrictive. In this work we use a

less restrictive cut based upon the γ-ray simulations, such that 95% of the simulated events that trigger the detector are accepted. Table 3 shows the significance and rate obtained from the 31 Crab Nebula pairs using the 95% cut for both the Neural Network and Kernel techniques. Referring to Table 1,

Technique	Significance	Rate
Neural Network	10.59σ	$4.49 \, \gamma \, \mathrm{min}^{-1}$
Kernel	11.60σ	$4.28 \, \gamma \, \mathrm{min}^{-1}$

Table 3. Results of the Crab Nebula ON/OFF data at the 95% cut used for spectral analysis

it is clear that this looser cut has almost doubled the accepted γ-ray rate. Although the significance has been severely reduced by the same cut, it is still appreciably high.

Fig. 1 shows the differential spectra of the Crab Nebula, obtained using the Neural Network and Kernel methods with the 95% cut. Table 4 lists the coefficients of the least-squares power law fits as indicated on both plots along with the results derived previously by Hillas et al. (1998).

Technique	Spectral Index (γ)	Flux Constant (α) ($\mathrm{m}^{-2} \, \mathrm{s}^{-1} \, \mathrm{TeV}^{-1}$)
Neural Network	2.29 ± 0.18	$(2.30 \pm 0.36) \times 10^{-7}$
Kernel	2.31 ± 0.15	$(2.31 \pm 0.32) \times 10^{-7}$
Hillas et al. (1998)	2.49 ± 0.06	$(3.20 \pm 0.17) \times 10^{-7}$

Table 4. Differential Spectrum of the Crab Nebula – Quoted errors are statistical only

5 Conclusion

The results in Tables 1 and 2 show that the Kernel technique enables sensitive discrimination between γ-rays and background events. However, both these sources are very strong emitters of γ-rays. The discovery of weaker γ-ray sources is at the heart of much current research in TeV γ-ray astronomy. For the Kernel technique to become a standard tool for γ-ray astrophysics, it must be demonstrated that it also outperforms the standard techniques for sources on the threshold of detectability. The paucity of confirmed TeV sources makes this a non-trivial task.

The energy spectra of the Crab Nebula derived by the Kernel spectral technique are in close agreement with the independent Neural Network analysis and also with previous results. All spectral techniques rely heavily on γ-ray simulations, and their accuracy is thus closely tied to the accuracy and breadth of the simulations. The range of energies simulated clearly constrains the calculated energies of real showers. Therefore, flux values calculated for energies close to the energy limits of the simulations must be treated with caution. The Kernel spectral analysis requires a correction factor to extract the known input spectrum from the γ-ray simulations. While this is not unusual in spectral analyses, it emphasises the inherent empiricism of such techniques.

References

Dunlea, S., Ph.D. Thesis (National University of Ireland), 2001.

Hand, D. J., Kernel Discriminant Analysis, John Wiley & Sons, Inc., Chichester, 1982.

Hillas, A. M., et al., ApJ 503, 744, 1998.

Holmström, L. and Sain, S. R., Technometrics 39, 91, 1997.

Kertzman, M. P. and Sembroski, G. S., NIM A 343, 629, 1994.

Mohanty, G., et al., Astroparticle Physics 9, 15, 1998.

Moriarty, P., et al., Astroparticle Physics 7, 512, 1997.

Moriarty, P. and Samuelson, F. W., GeV-TeV Gamma Ray Astrophysics Workshop: Towards a Major Atmospheric Cherenkov Detector VI, 338, 2000.

Punch, M., et al., Proc. 22nd ICRC (Dublin) 1, 464, 1991.

Samuelson, F. W., Ph.D. Thesis (Iowa State University), 1999.

Scott, D. W., Multivariate Density Estimation, p. 125, John Wiley & Sons, Inc., New York, 1992.
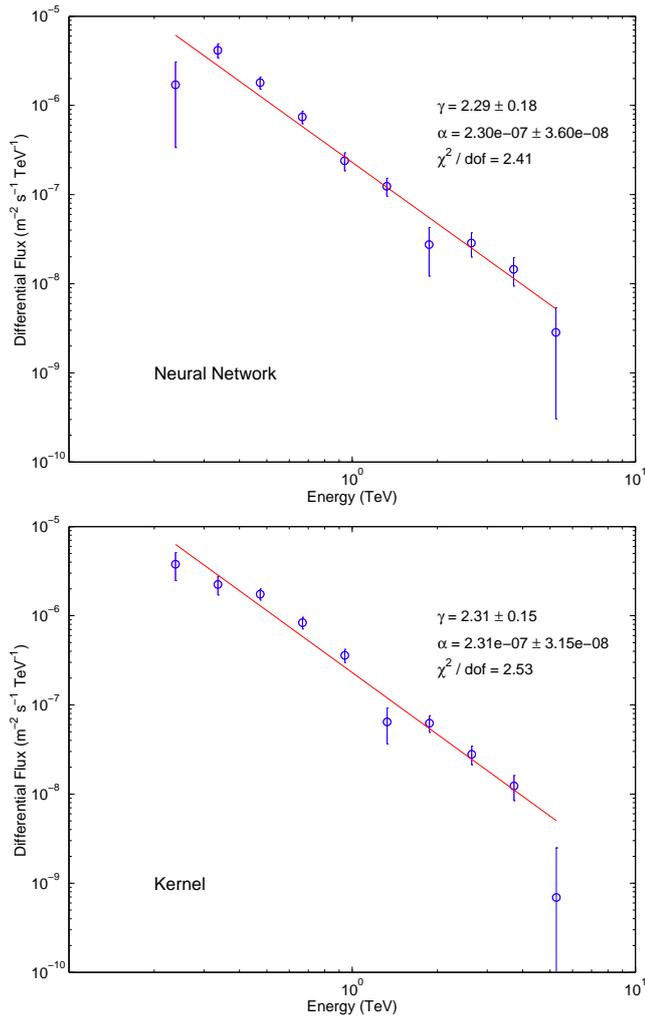
Fig. 1. *Crab Spectrum* – Differential Spectrum of the Crab Nebula determined with the Neural Network and Kernel techniques
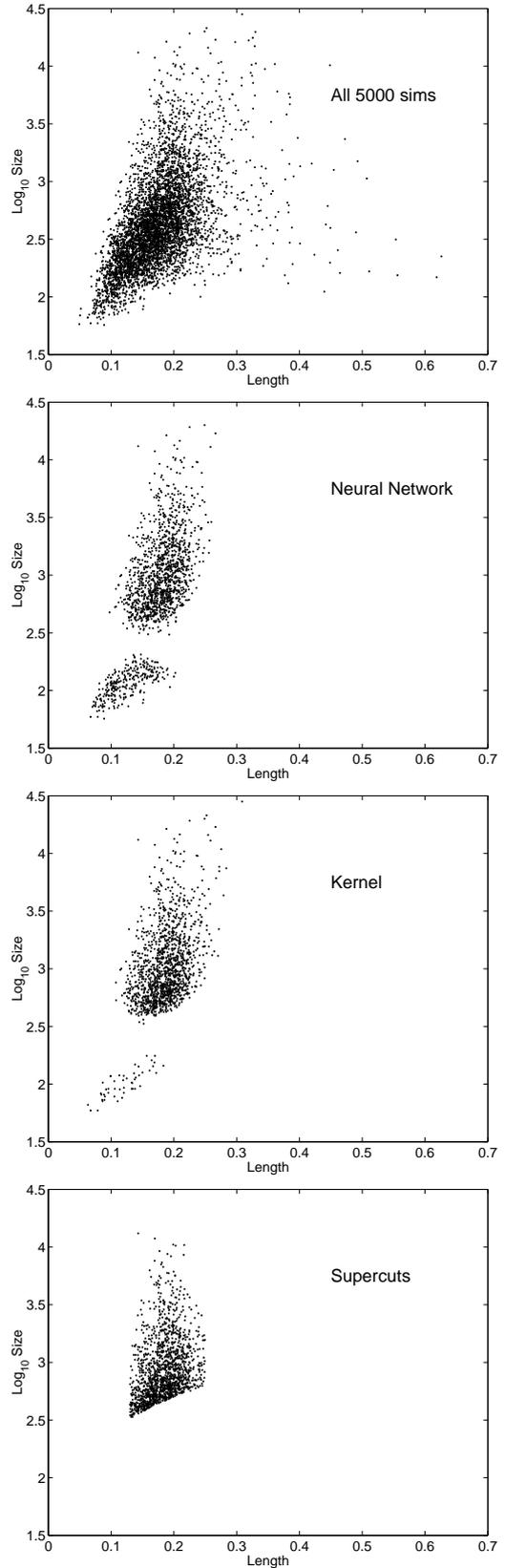

Fig. 2. *length* vs. $\log(size)$ – Distribution of *simulated γ-rays* passing the optimised cut in the 2-dimensional parameter space defined by *length* and $\log(size)$